

# **TOWARDS AN UNDERSTANDING OF THE RELATIONSHIP BETWEEN SPATIAL PROCESSING ABILITY AND NUMERICAL AND MATHEMATICAL COGNITION**

EDITED BY: Sharlene D. Newman and Firat Soylu  
PUBLISHED IN: Frontiers in Psychology





# 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 978-2-88963-534-4

DOI 10.3389/978-2-88963-534-4

## 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](mailto:researchtopics@frontiersin.org)

# TOWARDS AN UNDERSTANDING OF THE RELATIONSHIP BETWEEN SPATIAL PROCESSING ABILITY AND NUMERICAL AND MATHEMATICAL COGNITION

Topic Editors:

**Sharlene D. Newman**, Indiana University Bloomington, United States

**Firat Soylu**, University of Alabama, United States

**Citation:** Newman, S. D., Soylu, F., eds. (2020). Towards an Understanding of the Relationship Between Spatial Processing Ability and Numerical and Mathematical Cognition. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-534-4

# Table of Contents

- 04 Editorial: Towards an Understanding of the Relationship Between Spatial Processing Ability and Numerical and Mathematical Cognition**  
Firat Soylu and Sharlene D. Newman
- 06 Relationships Between Accuracy in Predicting Direction of Gravitational Vertical and Academic Performance and Physical Fitness in Schoolchildren**  
Wayne Haynes, Gordon Waddington, Roger Adams and Brice Isableu
- 22 Numerical and Non-numerical Predictors of First Graders' Number-Line Estimation Ability**  
Richard J. Daker and Ian M. Lyons
- 33 Spatial Presentations, but not Response Formats Influence Spatial-Numerical Associations in Adults**  
Ursula Fischer, Stefan Huber, Hans-Christoph Nuerk, Ulrike Cress and Korbinian Moeller
- 45 Sex Differences in the Performance of 7–12 Year Olds on a Mental Rotation Task and the Relation With Arithmetic Performance**  
Marleen van Tetering, Marthe van der Donk, Renate Helena Maria de Groot and Jelle Jolles
- 56 Mental Number Representations in 2D Space**  
Elena Sixtus, Jan Lonnemann, Martin H. Fischer and Karsten Werner
- 67 Mathematics Competence Level: The Contribution of Non-symbolic and Spatial Magnitude Comparison Skills**  
Marisol Cueli, Débora Areces, Ursina McCaskey, David Álvarez-García and Paloma González-Castro
- 76 The Importance of Ordinal Information in Interpreting Number/Letter Line Data**  
Christine Podwyssocki, Robert A. Reeve and Jason D. Forte
- 86 Individual Differences in Implicit and Explicit Spatial Processing of Fractions**  
Elizabeth Y. Toomarian, Rui Meng and Edward M. Hubbard
- 97 Multisensory Interactive Technologies for Primary Education: From Science to Technology**  
Gualtiero Volpe and Monica Gori
- 105 Building Blocks of Mathematical Learning: Virtual and Tangible Manipulatives Lead to Different Strategies in Number Composition**  
Ana Cristina Pires, Fernando González Perilli, Ewelina Bakata, Bruno Fleisher, Gustavo Sansone and Sebastián Marichal





# Editorial: Towards an Understanding of the Relationship Between Spatial Processing Ability and Numerical and Mathematical Cognition

Firat Soylu<sup>1\*</sup> and Sharlene D. Newman<sup>2†</sup>

<sup>1</sup> Educational Psychology Program, The University of Alabama, Tuscaloosa, AL, United States, <sup>2</sup> The Department of Psychology, The University of Alabama, Tuscaloosa, AL, United States

**Keywords:** numerical cognition, mathematical development, learning, spatial ability, mental number line, snarc, mental rotation

## Editorial on the Research Topic

### Towards an Understanding of the Relationship Between Spatial Processing Ability and Numerical and Mathematical Cognition

Engaging children in spatial thinking early may be important given that research as far back as Bingham's 1937 Aptitudes and Aptitude Testing reported that spatial thinking is associated with success in science, mathematics, and engineering fields. More specifically, recent research suggest a strong relationship between spatial ability and mathematics: Studies have found that performance in spatial tasks, for example mental rotation and visuospatial working memory, are correlated with mathematics achievement in school age children, that spatial representations are crucial in mathematical learning, and that spatial ability is closely related to numerical development. As a result, there have been recommendations to promote spatial play in preschool and incorporate spatial reasoning into elementary school curricula.

The mechanisms that underlie the relationship between spatial ability and mathematics are unclear. While there is some evidence to suggest that spatial ability can be trained, there are still questions concerning how best to train spatial ability. Further, how spatial representations for mathematical concepts play a role during the learning process is still under consideration. Therefore, this Research Topic presents 10 articles focusing on mechanisms underlying use of spatial representations in numerical cognition, the relation between spatial and mathematical abilities, and effectiveness of learning designs and interventions to foster mathematical performance, mediated by improvements in spatial abilities.

An important spatial concept that is used to represent numerical magnitudes is the *Mental Number Line (MNL)*. There are five studies in this Research Topic that focus on MNL, and the associated Spatial-Numerical Association of Response Codes (SNARC) and Numerical Congruity (NCE) effects. Daker and Lyons found that symbolic number comparison and non-verbal reasoning predicted a unique variance in a number-line estimation task, compared to approximate number processing ability, in first-grade children. They also found that the relation between symbolic number comparison and number-line ability was stronger for male students than female. Based on these results they argued for promoting children's understanding of symbolic rather than non-symbolic numerical magnitudes when learning from number-lines in the classroom. Fischer et al. found that while the different response modes (whole-body movements, hand movements, and verbal responses) did not affect the NCE and the SNARC effect, the presentation of a number line affected the two effects in opposite ways. While a larger SNARC effect was observed when

## OPEN ACCESS

### Edited and reviewed by:

Bernhard Hommel,  
Leiden University, Netherlands

### \*Correspondence:

Firat Soylu  
fsoylu@ua.edu

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

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

**Received:** 21 October 2019

**Accepted:** 06 January 2020

**Published:** 29 January 2020

### Citation:

Soylu F and Newman SD (2020)  
Editorial: Towards an Understanding  
of the Relationship Between Spatial  
Processing Ability and Numerical and  
Mathematical Cognition.  
Front. Psychol. 11:14.  
doi: 10.3389/fpsyg.2020.00014

a number line was presented, the NCE was only observed when no number line was presented, suggesting that the two effects reflect distinct processes. The SNARC effect is usually thought to reflect the existence of a horizontal mental number line, with values increasing from left to right. Some studies also show other alternative organizations of numbers in the two-dimensional space, in particular in a vertical fashion (“more is up”). Sixtus et al. investigated number representation over both vertical and horizontal dimensions within the same task-context. They found that the horizontal aspect of the spatial-numerical association did not have an effect on the behavioral performance, while the vertical one did, with higher performance when larger numbers were presented in the upper space. They concluded that numbers are conceptually associated with the vertical dimension when they are presented in a two-dimensional space. Podwysocki et al. investigated if the mental number line is unique to numbers or instead a general phenomenon that applies to any ordered list, by comparing performance with number and letter stimuli. They found comparable spatial-order effects for numbers and letters. They suggested that the mental line representation underlie ordered lists in general and is not unique to numbers, and the mental number line is supported by a common representation for all ordered sequences. Toomarian et al. investigated the relation between SNARC effect, fraction performance, and general mathematics ability. They replicated the SNARC effect with fractions and found that the individual SNARC effects correlated with performance on a fraction number-line estimation (NLE) task. Further, even though the NLE—but not SNARC—performance predicted scores in a fractions test and basic standardized mathematics performance, it did not predict algebra scores, implying that NLE may not be recruited for higher-order mathematical concepts.

In addition to spatial-numerical representations in numerical cognition, in particular the mental number line, there are other important aspects of the relation between spatial ability and mathematical thinking, some which were the foci for three further articles in this Research Topic. Haynes et al. reported that perception of vertical, a spatial orientation ability important to physical activity, is associated with male numeracy and female writing scores in children. Visual field independence (FI)—the degree of not being influenced by the background during perception of vertical—was found to be most strongly associated with numeracy, among other academic measures (i.e., reading, writing). Based on these results, the authors argued that systems for small visual angle spatial performance might be involved in number processing. van Tetering et al. presented results from a large scale study focusing on the relation between mental

rotation ability and mathematical performance with 7 to 12 year old children. They replicated the finding that mental rotation ability correlates with mathematical achievement, especially for boys. They also reported sex differences both in spatial ability and mathematics performance in the younger group (grades 2, 3, and 4) but not in the older one (grades 5 and 6). Based on these findings, the authors emphasized the importance of early experiences with spatial activities and play. Cueli et al. compared how non-symbolic and spatial magnitude comparison skills relate to mathematics ability in younger children. They reported that only the non-symbolic magnitude comparison scores predicted mathematics ability, pointing to the importance of training with non-symbolic magnitude tasks in early years of schooling.

Finally, another crucial aspect of the relation between spatial thinking and mathematical cognition is learning designs that involve spatial thinking and play, in order to promote mathematical learning and performance. Two studies in this Research Topic addressed this issue: Pires et al.’s intervention study with first-graders shows that technologically enhanced tangible manipulatives provide some advantages over explicitly virtual manipulatives. Children more readily interact with tangible manipulatives and the extent of interaction with manipulatives predict performance outcomes. Further, Volpe and Gori provide a set of principles on the use of multisensory technologies for learning and reflect on how multisensory technologies can be used in education.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

We would like to thank the authors for their valuable contributions and the reviewers for their constructive feedback.

**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 © 2020 Soylu and Newman. 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.



# Relationships Between Accuracy in Predicting Direction of Gravitational Vertical and Academic Performance and Physical Fitness in Schoolchildren

Wayne Haynes<sup>1\*</sup>, Gordon Waddington<sup>1</sup>, Roger Adams<sup>1</sup> and Brice Isableu<sup>2</sup>

<sup>1</sup> Research Institute for Sport and Exercise, Faculty of Health, University of Canberra, Canberra, ACT, Australia,

<sup>2</sup> Aix-Marseille Univ., PSYCLE, Aix-en-Provence, France

## OPEN ACCESS

### Edited by:

Firat Soylu,  
The University of Alabama,  
United States

### Reviewed by:

Marci E. Gluck,  
National Institute of Diabetes  
and Digestive and Kidney Diseases  
(NIDDK), United States  
Michael Barnett-Cowan,  
University of Waterloo, Canada

### \*Correspondence:

Wayne Haynes  
Wayne.Haynes@canberra.edu.au

### Specialty section:

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

**Received:** 11 March 2018

**Accepted:** 02 August 2018

**Published:** 23 August 2018

### Citation:

Haynes W, Waddington G, Adams R  
and Isableu B (2018) Relationships  
Between Accuracy in Predicting  
Direction of Gravitational Vertical  
and Academic Performance  
and Physical Fitness  
in Schoolchildren.  
Front. Psychol. 9:1528.  
doi: 10.3389/fpsyg.2018.01528

Enhanced levels of cardio-respiratory fitness (CRF) and physical activity (PA) are both positively associated with health and academic outcomes, but less is known about the spatial processing and perceptual components of PA. Perception of vertical (PV) is a spatial orientation ability that is important for PA, and is usually measured as relative accuracy in aligning an object to gravitational vertical against a tilted background. However, evidence is inconclusive regarding the relationship of PV to educational outcomes – most importantly, numeracy. Students were recruited from primary schools in the Australian Capital Territory. A group of 341 (females  $n = 162$ , mean age 11.3 years) children performed all the tests required for this study. A computerised rod and frame test of PV employing a small ( $20^\circ$ ) visual angle was administered, and socio-economic status (SES), national education test results (NAPLAN, 2010), and CRF and PA data were collected. Correlation and hierarchical regression analysis were used to examine the inter-relationships between PV and CRF, PA, SES and NAPLAN results. The two extreme quartile score groups from the measures of PV, PA and CRF were examined in relation to NAPLAN scores. PV scores arising from testing with a small visual angle and SES were found to be significantly associated with overall academic scores, and with the Numeracy, Reading, and Writing components of academic performance. Female gender was significantly associated with Writing score, and male with Numeracy score. Being less influenced by the background tilted frame, and therefore having visual field independence (FI), was associated with significantly higher academic scores, with the largest effect in Numeracy scores (effect size,  $d = 0.82$ ) and also associated with higher CRF and PA levels. FI was positively associated with all the academic modules examined, and most strongly with Numeracy test results, suggesting that FI provides an indicator of STEM ability. These findings suggest that further longitudinal research into strategies designed to enhance visual FI deserve consideration, with a focus on specialized PA programs for pre-pubescent children. It is possible that small visual angle spatial tasks during PA may stimulate neural networks involved in numerical cognition.

**Keywords:** perception of vertical, spatial processing ability, numerical cognition, STEM, academic performance, physical activity, cardio-respiratory fitness, pre-pubescent children

## INTRODUCTION

Enhanced levels of physical activity (PA) and cardio-respiratory fitness (CRF) in children are proposed to be positively associated with health, with wellness (Smith et al., 2014; Poitras et al., 2016) and also with academic performance (for a review, see Tomporowski et al., 2015; Ruiz-Ariza et al., 2017). While little controversy exists regarding the positive association of elevated levels of CRF and PA with enhanced child health and wellness outcomes (Smith et al., 2014; Poitras et al., 2016), some researchers have queried the extent to which PA and CRF are directly associated with academic performance, and have suggested the possible involvement of other factors (Pesce, 2012; Diamond and Ling, 2016). Diamond and Ling (2016) argue that PA undertaken with a focus on co-activating elements engaged in cognition provides the best possible combination of activities to enhance academic performance and cognitive function in children. Pesce (2012) has also suggested that while a significant portion of previous research has focused primarily on relationships between quantitative elements of PA with academic and cognitive ability, qualitative factors have been relatively neglected. Quantitative modes of PA relate exclusively to the duration and intensity of physical activities, whilst qualitative modes include more complex tasks requiring greater involvement of cognitive networks, by stimulating perceptual mechanisms and executive control pathways (Pesce, 2012; Tomporowski et al., 2015).

Spatial ability includes both a perceptual component and a qualitative factor associated with PA (Pesce, 2012). Spatial ability relates to the organization of sensory experiences in perception of physical spaces, incorporating embedded objects and the perceiver. Understanding the association between spatial perception and PA is important for pedagogical reasons, as evidence suggests that spatial abilities strongly predict analytical aptitude, involvement and performance in Science, Technology, Engineering and Maths (STEM) education (Shea et al., 2001; Wai et al., 2009; Ganley et al., 2014). Recently, a working paper by the OECD suggested that spatial ability and STEM learning are associated, and that spatial abilities are malleable (Newcombe, 2017).

An individual's ability to estimate the direction of gravitational vertical, known as subjective visual vertical (SVV) is an important spatial ability when selecting a frame of reference to be used in the organization of upright orientation and postural alignment, and it requires both sensorimotor and cognitive participation (Isableu et al., 2010; Agathos et al., 2015). To successfully engage in any PA, an individual must select a suitable sensory frame of reference for spatial orientation so as to functionally align their body axis with primary vertical and horizontal axes, by means of information collected from vestibular and somato-sensory receptors and/or from visual examination of the physical features of their environment (Isableu et al., 2010; Agathos et al., 2015).

VV predictions begin with the engagement of the vestibular system providing a head centered frame of reference for upright orientation and alignment (Tarnutzer et al., 2009). Specifically, the saccule otolith is the primary sensory apparatus for detecting linear accelerations and head tilts in the vertical plane, with the

signal disambiguated by integration of input from the posterior semi-circular canal detecting angular acceleration in the frontal plane. These integrated signals are then combined with roll plane proprioceptive inputs from the cervical spine enabling the head to be isolated as the segment exploited to form the SVV-based head stabilized in space strategy (Assaiante and Amblard, 1993; Tarnutzer et al., 2009; Schuler et al., 2010; Clemens et al., 2011; Medendorp and Selen, 2017). The central nervous system, using probabilistic methods, integrates and re-weights these multi-sensory inputs from multiple sources to formulate the most reliable spatial frame of reference and thereby align the body to gravitational vertical (Carver et al., 2006). Recent theory also holds that the central nervous system can internally construct a prediction of the direction of gravitational vertical, based on past experiences, when sensory information is ambiguous, unreliable or inaccessible (Mergner et al., 2003; Carver et al., 2006; MacNeilage et al., 2007; Barra et al., 2010).

Many spatial tasks require the central nervous system to estimate physical spatial quantities (kinematic or dynamic) and this is the basis for the theory of dimensional magnitudes as a foundation for numeracy (Walsh, 2003). For example, dimensional magnitudes include displacement and velocity, directions and orientation in three dimensional Euclidean space with force, mass or inertia related to a target or to the body. These physical properties are bound to dimensions and quantities and are embodied in their cognitive relationships (Kozioł et al., 2012). Further, these physical properties are estimated relative to a spatial frame of reference, with the foundational one being SVV. Therefore to estimate distance, dimensions or other physical properties immersed in the spatial world, SVV predictions underlie their cognitive representation. This may then serve as a scaffold for spatial cognition and its association with STEM subjects. Support for this proposition also comes from Verdine et al. (2017), stating that the foundation of mental models representing numbers and magnitudes may be in spatial representations.

SVV is typically measured using the rod and frame test (Witkin and Asch, 1948; Oltman, 1968; Isableu et al., 2010). In all versions of the rod and frame test, an individual is positioned in a darkened environment and exposed to a tilted illuminated frame encasing a movable, illuminated rod, tilted from the vertical. Individuals are tasked with aligning the rod to gravitational vertical, with error in doing so thus providing a measure of SVV.

Two primary strategies exist in the formulation of gravitational vertical estimations in humans. First, the tilted frame may strongly affect ability to align the rod to vertical. Those affected typically use a visually dominant strategy that is heavily reliant on sensory inputs from peripheral vision (the frame), leading to biased estimation of vertical (Isableu et al., 2010) and these individuals are classified as visually field dependent. Evidence suggests that field dependent participants experience difficulties in up-weighting internal postural signals (vestibular and somatosensory) for obtaining more accurate perceptual constructs (Isableu et al., 2010; Agathos et al., 2015). Further, Agathos et al. (2015), provide evidence that field independent participants possess higher selective attentional control, more stabilized eye movements and greater inhibitory



capabilities than field dependent subjects, with significant advantages in academic tasks, particularly STEM performance.

Secondly, individuals whose perception of vertical (PV) is more independent of the visual frame are said to use a field independent strategy. Field independent individuals most often employ a flexibly arranged and re-weighted integration of vestibular cues and somatosensory inputs with vision for PV (Isableu et al., 2010). This strategy provides a task-specific prediction of gravitation vertical which reliably provides lower error in the estimate of vertical than does the field dependent strategy, and applies even when confronted by ambiguous, conflicting or challenging visual environments (Kent-Davis and Cochran, 1989; Isableu et al., 2010). Finally, with extensive experience of sensory signals related to vertical alignment, predictive mechanisms are refined leading to an internal prediction of the direction of gravitational vertical (Lopez et al., 2011) that can act as a critical arbiter in sensorily deprived environments (Mergner et al., 2003; Barra et al., 2010).

Early researchers into the field independent and field dependent phenomenon suggested the rod and frame test and a test of spatial cognition – the Embedded Figures Test – to examine similar properties of spatial ability, and these were at first used for field dependent and field independent classification. However upon further review, correlation analysis found that, whilst both measures of field independent and field dependent share similar characteristics, they do not measure the same thing. For example from over 300 studies, Arbuthnot (1972) found the correlation between the two measures to be moderate to strong but indicating that they were not inter-changeable. It is a widely held view that the Embedded Figures Test is a more effective test of spatial cognition in that it reflects a child's ability to perform numeracy tasks. Classification of field dependent and field independent should stipulate the type of measure used in assessment (Arbuthnot, 1972).

Previous longitudinal research examining field dependence and independence in young children through to adolescence indicates that there is progress from being relatively strongly field dependent to being relatively more field independent by adulthood; with boys more field independent than girls (Witkin et al., 1967; Bagust et al., 2013). At an individual level, a child from about 7 years can be classified as relatively field independent or field dependent compared to a cohort of similarly aged children. As the child in this group ages, they will generally become more field independent but their ranking relative to other subjects in this group will not significantly change (Witkin et al., 1967).

Both early research and more contemporary studies provide evidence of high inter-individual variability in the accuracy of predicting vertical using the rod and frame test, whilst within the individual there is a stable preference for the style or mode of vertical perception, leading to self-consistency in performance on a number of spatial tasks (Witkin, 1959; Isableu et al., 2010). This prompted early researchers to identify PV as a perceptual style, whereby individuals are ranked according to their high or low error rate on performance of tasks measuring vertical perception, and then grouped as either field dependent or field independent (Witkin, 1959). Conventionally, a group of individuals performing the rod and frame test is divided into top

and bottom quartiles, based on their rod and frame test results (Isableu et al., 1997). The low perceptual error quartile is classified as field independent and the highest perceptual error quartile classified as field dependent.

Field independent and field dependent individuals appear to analyze information differently, using different cognitive strategies and have different levels of cognitive neural complexity in problem solving, with field independent using an analytic strategy by breaking down the complex structure of a stimulus, and field dependent employing a more holistic style (Witkin and Goodenough, 1981; Davis and Cochran, 1982; Jia et al., 2014; Agathos et al., 2015). Field independent individuals are also described as having an articulated body concept and percept, in which they view the available sensory array as discrete and have the ability to cognitively restructure available information to formulate more accurate perceptual predictions (Witkin, 1959; Isableu et al., 2010; Agathos et al., 2015).

Expanding on the relationship between cognitive functions and rod and frame test performance in measuring SVV, research has consistently revealed differences in general learning and memory between field dependent and field independent individuals, as determined by the rod and frame test (Gough and Olton, 1972; Goodenough, 1976; Blowers and O'Connor, 1978; Shinar et al., 1978; Berger and Goldberger, 1979; Amador-Campos and Kirchner-Nebot, 1999). Evidence from the Wechsler Intelligence Scale for Children supports the view that field independent children perform with greater effectiveness in tests of intelligence (Goodenough and Karp, 1961; Dreyer et al., 1971).

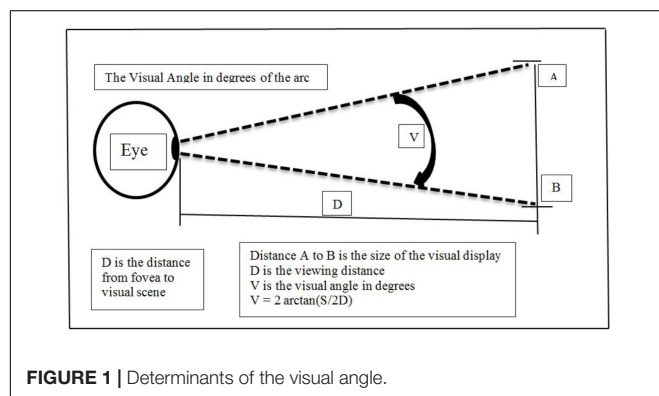
Whilst evidence supports an association between rod and frame test performance and cognition, controversy still exists regarding the association with academic performance in pre-adolescent children (**Table 1**). Support for the argument that the rod and frame test is related to elements of academic performance comes from Kagan and Zahn (1975), Kagan et al. (1977) and Canavan (1969), all of whom used a mechanical version of the RFT called the "Man in the Frame" rod and frame test, with a smaller visual angle of 22 degrees. Alternatively, several studies using the traditional mechanical rod and frame test with a larger visual angle of 28 degrees (Oltman, 1968) have found no significant relationship between rod and frame test results and measures of academic performance (Buriel, 1978; Allan et al., 1982; Wong, 1982; Tinajero and Páramo, 1997).

The reported associations between SVV, PA and motor coordination are unambiguous. Field independent compared to field dependent individuals have higher levels of PA and sports participation, and are seen as possessing greater sports potential (Liu and Chepyator-Thomson, 2008, 2009), more advanced motor competency and motor coordination (Meek and Skubic, 1971; Golomer et al., 1999) and greater ability to learn novel motor tasks (Hodgson et al., 2010). Finally, more athletes are field independent than non-athletes (Brady, 1995). Most of the studies reviewed examined SVV using the mechanical rod and frame test with a visual angle of 28 degrees (Oltman, 1968). The rod and frame test appears to be able to distinguish between athletes and non-athletes due to the higher engagement of vestibular and proprioceptive sensors (Raviv and Nabel, 1990; Liu and Chepyator-Thomson, 2009).

**TABLE 1 |** Summary of childhood studies into the association of between rod and frame tests measuring subjective visual vertical using either a small or large visual angle with academic performance.

Study	Study methodology				Academic association			
	Age	N	Test	Visual angle	Research	Results association with rod and frame test	Numeracy	Reading English
Kagan et al., 1977	10–12	230	MF	22° small	Correlation	Maths and English moderate to strong correlation	✓	✓
Kagan and Zahn, 1975	8–12	134	MF	22° Small	Correlation and multiple regression	Sig correlation maths ( $\beta = -0.22$ ) and English ( $\beta = -0.19$ )	✓	✓
Canavan, 1969	10–12	1167	MF	22° small	Correlation	Maths and English moderate to strong correlation	✓	✓
Buriel, 1978	7–10	80	Hybrid large visual angle, MF RFT design	28° large	Correlation ANOVA	No relationship maths or English	×	×
Wong, 1982	10–11	90	PRFT	28° large	Correlation regression	No relationship maths, English or language	×	×
Tinajero and Páramo, 1997	13–16	408	PRFT	28° large	MANOVA	No significant source of variance languages, science, maths, English and overall	×	×

Methodology: man in the frame (MF) RFT with visual angle 22°, portable rod and frame test (PRFT) with visual angle 28° Hybrid – portable rod and frame test with the man in the frame visual task.

**FIGURE 1 |** Determinants of the visual angle.

The visual angle projected onto the eye during the rod and frame test is an important factor in SVV performance. The visual angle is produced by two straight lines drawn from the peripheral points of a seen object to the fovea of the eye and is normally stated in degrees of arc (**Figure 1**), and visual angles are an important discriminating characteristic between small and large field of view environments in spatial perception (Wang et al., 2014).

Large-scale field-of-view environments naturally produce large visual (retinal) angles. When the participant is an active part of the environment the PV incorporates the involvement of proprioceptive and vestibular input, causing visual-vestibular interactions generated by the head, body and eye movements that are made in order to take in the full visual array. Research suggests that large visual angles promote perceptions related to environmental interactions, such as movement and navigation,

and may be more closely associated with PA measures (Quaiser-Pohl et al., 2004). On the other hand, perceptions formed by small scene environments characteristically employ small visual angles with reduced eye, head or body movements and are more reliant on foveal-based ventral visual streams. Quaiser-Pohl et al. (2004) provide evidence that small visual angle RFT scores have a moderate to strong correlation with other small-scale environment spatial cognitive tasks, the Water-Level-Task and a Mental-Rotations-Test, in 10–12 years old children, while large scale environments with large visual angles were uncorrelated with the small visual angle RFT. Quaiser-Pohl et al. (2004) suggested that it is the small-scale visual environments that tap directly into the spatial reasoning mechanisms associated with STEM subjects.

Finally, SES was included as it has been found to be a strong predictor of academic performance (Mezzacappa, 2004; Sirin, 2005) and CRF (Duncan et al., 2002) and has previously been associated with performance with the rod and frame test (Maceachron and Gruenfeld, 1978). Therefore, including SES as a benchmarking measure enables the assessment of the relative importance of different associations.

The current study examined the cross-sectional relationships between PA, CRF, SES, academic performance and a Computerised Rod And Frame Test (CRAFT) in 10-year-old children. This study is the first to examine PV using a small visual angle CRAFT in association with both academic and physical health-related measures in a large cohort of 10 years old children.

The primary hypothesis was that measurement of SVV using a small visual angle would be significantly associated with measures of academic performance, with the greatest

association arising in numeracy scores and would generate significant effects when comparing field dependent with field independent individuals. It was further hypothesized that the SVV results would have significantly smaller associations with PA-related measures, in a manner different from findings in previous studies using a large visual angle task. Finally, it was proposed that grouping participants into the highest and lowest PA and CRF quartiles would show that children who exhibit higher CRF and PA levels have a significant advantage in areas of academic performance and support the role of enhanced PA levels as a moderating effect on academic performance.

## MATERIALS AND METHODS

### Participants

The data were collected as part of the Lifestyle of our Kids (LOOK) that included 853 (418 females) children with mean age 11.3 years [(SD 0.3) years] in a longitudinal study that involved 29 elementary (primary) schools in Canberra in the Australian Capital Territory (Telford et al., 2009). From the total LOOK sample there were 341 (162 females) children who performed all the tests required for the current study. The numbers of children in the regression analyses on academic score dependent variables were; Overall Academic Score ( $n = 341$ ), Numeracy ( $n = 345$ ), Reading ( $n = 345$ ) and Writing ( $n = 346$ ). The difference in the numbers of children in this study compared to the total involved in the research was primarily related to access to children on testing days. Elements of testing were conducted at different times, some at school and some at a hospital location, and if children did not attend school on the relevant testing days, their data was not collected. The measure of SES used in study was obtained from a parent questionnaire on the level of education achieved, however not all parents returned this questionnaire. Further, some children did not sit the Australian National Assessment Program Literacy and Numeracy (NAPLAN) exam due to parental choice. For the PA measure, the children were encouraged to wear the pedometer each week, however not all did so.

Academic tests were conducted in the same year as with the other measures, except for the CRAFT with data collected 4 months before the academic tests. The relative uniformity of the schools in the study reflected the fact that all schools were part of a local public education system, receiving similar funding. This study was approved by the Australian Capital Territory Health and Community Care Human Research Ethics Committee. Participation by the children was voluntary and informed consent for involvement was received from parents or guardians.

### Measures

#### Perception of Vertical Measurement

With the CRAFT, a 430 mm flat screen monitor presented an image of a tilted illuminated frame enclosing a tilted illuminated rod. The square illuminated frame was 185 mm wide and

surrounded by a blackened cylinder 330 mm diameter and 500 mm long (Haynes et al., 2008). The viewing tube was set 30 mm from the computer screen. The computer screen, viewing tube and child were all covered by a dark cloth eliminating any external light source. When tested, the child was seated with the viewing tube at eye level and their chin placed on a foam rest within the entrance of the viewing tube, and to limit proprioceptive and vestibular inputs during the trials the child was requested to remain still. The visual angle of twenty degrees was calculated from the distance of the chin rest on the viewing tube to the square frame appearing on the computer screen. Designing the CRAFT with a small visual angle and reducing head motion enabled measurement of a child's prediction of gravitational vertical with reduced visual-vestibular interactions and proprioceptive inputs.

Children were asked to imagine that the rod was a rocket ship aimed to shoot straight up to an imaginary moon. The children were told that in the event the rocket ship was not pointing straight up, then it would crash. The child was given two practice trials to "launch the rocket ship to the moon." If performance was poor, the child was shown the error and instructed on the correct alignment.

The test consisted of ten trials presented in random order, consisting of five frames tilted clockwise and five frames tilted anti-clockwise at eighteen degrees to vertical. The rod was positioned either twenty degrees positive or negative to vertical with the frame tilted eighteen degrees clockwise or counter-clockwise (Figure 2). The child could move the rod using a handheld mouse and when satisfied with the "rocket ship" alignment, pressed the space bar to record their response, thereby giving an angle from the vertical, and were then automatically moved onto the next trial. The mean of the ten absolute (unsigned) errors for each subject was calculated. The reliability of the mechanical rod and frame test has been previously found to be acceptable (Witkin et al., 1967).

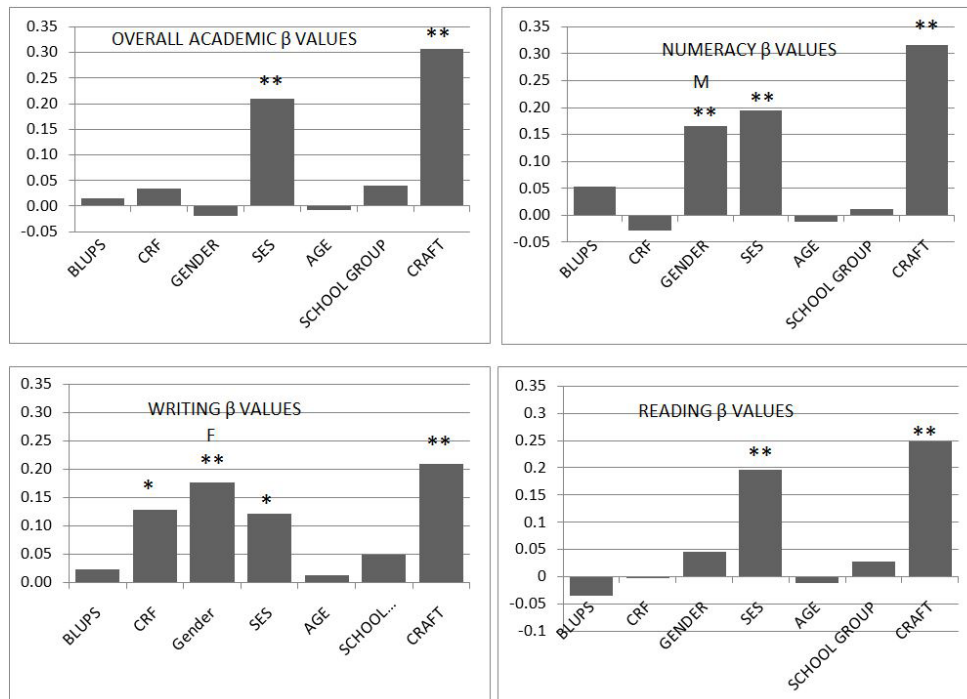
#### Socio-Economic Status Measurement

The measure of SES employed was obtained from a questionnaire enquiring about the level of educational attainment by either of the parents. Score of one indicated year 10 level, score two was high school level of education and score three was tertiary qualification. Because SES has been found to be a strong predictor of academic performance (Sirin, 2005), knowing SES is critical in any study of academic performance differences.

#### Physical Activity and Physical Fitness Measurements

Cardio-respiratory fitness was assessed by a twenty-meter multistage run (Tomkinson et al., 2003) using methods described by Telford et al. (2009). Participants ran between two lines, 20 m apart, while keeping pace with a loud beeping sound arising from a sound amplifier. The measure of CRF was the number of stages reached in the multistage run.

Physical activity was measured by AT pedometers (New-Lifestyles, Lee's Summit, MO, United States) considered to be sufficiently valid and reliable (Beets et al., 2005) as described by Telford et al. (2009). PA was measured by pedometer use for



**FIGURE 2 |** Beta weights from the hierarchical regressions. Dependent variables are Overall academic scores, Numeracy, Reading, and Writing. Independent variables are Physical Activity (PA), Cardio-respiratory Fitness (CRF), Socio-Economic Status (SES) and Computer Rod And Frame Test (CRAFT). CRAFT produced negative beta values because low error scores are associated with higher scores on academic variables, however for comparison with the other independent variables, the CRAFT beta values are shown here as positive. Gender beta values are shown as positive for the better performing gender – MALE (M) and FEMALE (F). No changes were made to the signs of the beta values for PA, CRF, SES. Significance: \* $P \leq 0.05$ , \*\* $P \leq 0.01$ .

seven consecutive days by every child, conducted as described in Telford et al. (2009). A PA index was formulated engaging the Best Linear Unbiased Predictor (BLUPS) (Robinson, 1991).

### Academic Performance Measurement

Academic performance was measured by the NAPLAN tests conducted in year five<sup>1</sup>. The Australian Department of Education is responsible for all testing and scaling of results collected from all Australian primary school children. Numeracy, Reading, and Writing scores were collected from the NAPLAN results for the current project. Persuasive and narrative writing skills were examined in the Writing task. Literacy proficiency is tested by reading tasks, and focuses on the reading and comprehension of written English. The Year five reading text included biographies, autobiographies and persuasive passages. Several competencies were examined in the Year five NAPLAN numeracy test, including algebra, quantities, patterns and measurements and space concepts.

### Statistical Analysis

Descriptive statistics were obtained first. Bivariate correlations between the procedures were calculated as a validity check, to determine if the individual variables within a class had significant relationships. Correlation analysis used the maximum number of

participants available so as to get the best estimate of descriptive statistics. Effect size for bivariate correlations using Pearson's coefficient of correlation was determined using the classification provided by Cohen (1992), with a small effect at 0.1, medium at 0.3 and large at 0.5.

To examine the contribution of SVV in the explanation of academic performance in ten year old children, a hierarchical multiple regression analysis was performed. A series of three-step hierarchical linear regressions were undertaken on dependent variables – Year five NAPLAN results in Reading, Writing, Numeracy and overall academic scores. Independent predictor variables were: CRF, PA, CRAFT, and SES. As well, sex (male, female) was included as a dichotomous variable in the hierarchical linear regression model. All statistical assumptions were met in the analyses. In step one, independent variables PA and CRF were regressed. Sociocultural factors – SES and the dichotomous gender variable were included in step two. The third step involved including the measure of SVV measured with the CRAFT. In order to adjust significance level for multiple comparisons, the Bonferroni correction was employed and the alpha level was set at 0.01.

The upper and lower quartiles were determined for all independent variables. The two groups were defined as “first quartile” (children who achieve the highest competency in a task) and “fourth quartile” (children who have the lowest competency in a task).

<sup>1</sup><https://www.nap.edu.au/>



Students in the first quartile (low CRAFT error scores indicating an ability to accurately assess gravitational vertical) were classified as field independent; and fourth quartile participants (high error in predicting vertical) and were classified as field dependent. Quartile splits for PA and CRF were also applied to the academic variables. To assess the influence of classification of individuals as field independent or field dependent, independent groups *t*-tests were applied to the quartiles produced from the CRAFT, first with dependent variables Numeracy, Reading, Writing and overall academic performance, as well as PA CRF, and SES. Independent-groups *t*-tests were also performed on the quartile splits on PA and CRF scores with dependent measures; Reading, Writing, Numeracy and overall academic score, and with CRAFT performance. Means and standard deviations were calculated for Reading, Writing, Numeracy and overall academic scores to determine overall academic performance in the different groups. Using the quartile mean scores and standard deviation, effect size was calculated using Cohen's *d*, the standardized mean difference, with a small effect designated as 0.2, medium as 0.5 and large as 0.8 (Cohen, 1992). Statistical analyses were performed using SPSS version 21 with statistical significance set at 0.05.

## RESULTS

**Table 2** provides descriptive statistics. Bivariate correlations between the three measures of academic performance from the national examination scheme (reading, writing, and numeracy) and summed academic scores, together with the CRAFT, CRF, PA, and SES results are presented in **Table 3**.

Small to medium-sized, statistically significant correlations were observed between the CRAFT and individual measures of academic performance, measured by the NAPLAN tests in ten year old school children for reading ( $r = -0.24$ ,  $p < 0.001$ ), writing ( $r = -0.2$ ,  $p < 0.001$ ), numeracy ( $r = -0.35$ ,  $p < 0.001$ ) and summed academic scores ( $r = -0.31$ ,  $p < 0.001$ ). Small to medium sized correlations were found between SES and academic performance in reading ( $r = 0.24$ ,  $p < 0.001$ ), writing ( $r = 0.21$ ,  $p < 0.001$ ), numeracy ( $r = 0.25$ ,  $p < 0.001$ ) and overall academic scores ( $r = 0.27$ ,  $p < 0.001$ ).

Cardio-respiratory fitness association with numeracy was significant and classified as small ( $r = 0.12$ ,  $p = 0.007$ ), as it was with Overall Academic Scores ( $r = 0.12$ ,  $p = 0.009$ ), and with writing having a small to medium effect size ( $r = 0.15$ ,  $p < 0.001$ ). No association was discovered between CRF and reading. PA had a small to medium, significant relationship with numeracy ( $r = 0.15$ ,  $p < 0.001$ ) only.

Socio-economic status had no significant correlation with CRF or PA. SES produce a small significant association with CRAFT ( $r = -0.11$ ,  $p = 0.021$ ). Moderate and significant correlations were observed between PA and CRF ( $r = 0.49$ ,  $p < 0.001$ ). Moderate to large and significant correlations were found between numeracy and reading ( $r = 0.69$ ,  $p < 0.001$ ), numeracy and writing ( $r = 0.49$ ,  $p < 0.001$ ) and reading and writing ( $r = 0.58$ ,  $p < 0.001$ ). PA and CRF correlation scores with the CRAFT were not significant (all  $p > 0.1$ ).

Hierarchical linear regression analysis is exhibited in **Table 4**. Beta values were tabulated from the regression analysis in individual column charts (**Figure 1**) for dependent variables; Overall academic scores, numeracy, reading and writing with independent variables being PA, CRF, SES, and CRAFT score. In relation to overall academic scores, hierarchical linear regression analysis results from step one revealed no significant relationships with PA and CRF variables. Step two was significant ( $R^2 = 0.09$ ) and provides evidence of a significant association with overall academic scores arising from SES. No significant association found from the gender variable. Step three including only the CRAFT variable with scores tending to be higher (greater field dependent) for poorer academic scores, giving a negative B-weight, and PV variable alone accounted for 8.4% of the summed overall academic scores variance. CRAFT measurement performance ( $\beta = -0.3$ ;  $t = -5.85$ ,  $p < 0.001$ ) had significantly higher beta values than SES measures ( $\beta = 0.26$ ;  $t = 5.22$ ,  $p < 0.001$ ). The model accounted for 16.8% of the variance.

In explanation of Numeracy, hierarchical multiple regressions revealed that from step one, neither CRF nor PA contributed significantly to the variance of the model. Introducing step two explains an additional 11.4% of variance with SES and male gender both significant. Step three provides evidence of 19.9% of variance in the model (adj  $R^2 = 0.2$ ) with the CRAFT variable explaining 7.9% of the variance of the model. Significant beta values in step three were PV score (CRAFT) ( $\beta = -0.288$ ;  $t = -5.821$ ,  $p < 0.001$ ), SES ( $\beta = 0.22$ ;  $t = 4.51$ ,  $p < 0.001$ ) and sex (male) ( $\beta = 0.21$ ;  $t = 3.895$ ,  $p < 0.001$ ).

The results of step one, when examining the Reading dependent variable, reveal PA and CRF variables produce no significant association. In step two, SES and gender dichotomous variable were entered, with only SES explaining variance in the model ( $R^2 = 0.59$ ). When all five variables were included in the step three model, only SES ( $\beta = 0.254$ ;  $4.984$ ;  $p < 0.001$ ) and SVV ( $\beta = -0.239$ ;  $t = -4.611$ ;  $p < 0.001$ ) variables were significant. SVV variable measured by the CRAFT explained 4.1% of the variance. Together the six independent variables accounted for 10.1% of the variance in Reading scores.

To examine the unique contribution of SVV in the explanation of writing skill in eleven year old children, a three step hierarchical multiple regression analysis was performed. Step one examining PA and CRF variables provided evidence of a significant small association arising in CRF. 1.4% of the variance arising in the model was explained in step one. By step two a further significant 8% of the variance is explained in the model with SES and gender (female) significant in their contribution. Step three, including the five variables in the model, provides 12.4% of the variance explained by the model, with SVV explaining 5.4% of the model. The three independent variables significantly related to the results in this model were SES ( $\beta = 0.19$ ;  $t = 3.6$ ;  $p < 0.001$ ), SVV ( $\beta = -0.21$ ;  $t = -3.98$ ;  $p < 0.001$ ) and gender (female) ( $\beta = -0.15$ ;  $t = -2.76$ ;  $p = 0.006$ ).

To compare the effects of classification regarding a field independent or field dependent perceptual style, with the quartiles produced from the CRAFT, *t*-tests were applied to examine overall academic scores, numeracy, reading and writing (**Table 5**). Numbers of participants classified as field dependent

**TABLE 2 |** Descriptive statistics.

Independent variable	N	Min	Max	Mean	St. Dev
CRAFT	541	0.8	55	5.6	5.7
Overall academic scores	705	202.33	716.7	489.7	62.8
Numeracy	714	307.00	771	484.8	71.5
Writing	716	89.00	692	483.5	71.7
Reading	715	98.00	754	499.5	77.9
SES	593	1	3	2.49	0.72
PA	853	64.99	133.7	100.1	11.1
CRF	534	1.30	11.5	5.34	1.9
Valid (listwise)	341				

Sample means, standard deviation (St. Dev) minimum (min) and maximum (max). Computerised Rod And Frame Test (CRAFT), Socio-Economic Status (SES), Cardio-Respiratory Fitness (CRF), and Physical Activity (PA). SES score represents the highest educational attainment by a parent with one representing lowest level of education with three representing a university degree or similar. PA score represent an index of PA with a low score indicating low levels of PA and high scores representing high levels of PA CRF represented by a shuttle run score with the number of stages completed.

**TABLE 3 |** Bivariate correlation coefficient matrix.

		1	2	3	4	5	6	7	8
1. CRAFT	r	1	−0.31**	−0.345**	−0.2**	−0.243**	−0.111*	−0.078	−0.056
	Sig.		0.001	0.001	0.001	0.001	0.021	0.07	0.239
	N	541	507	513	514	514	429	541	442
2. OAS	r		1	0.852**	0.806**	0.896**	0.274**	0.074	0.116**
	Sig.			0.001	0.001	0.001	0.001	0.051	0.009
	N		705	705	705	705	526	705	503
3. NUMERACY	r			1	0.493**	0.691**	0.254**	0.146**	0.12**
	Sig.				0.001	0.001	0.001	0.001	0.007
	N			714	708	709	532	714	510
4. WRITING	r				1	0.582**	0.209**	0.051	0.146**
	Sig.					0.001	0.001	0.173	0.001
	N				716	711	532	716	512
5. READING	r					1	0.24**	−0.007	0.034
	Sig.						0.001	0.846	0.449
	N					715	533	715	510
6. SES	r						1	−0.01	0.032
	Sig.							0.806	0.52
	N						593	593	414
7. PA	r							1	0.487**
	Sig.								0.001
	N							853	534
8. CRF	r								1
	Sig.								
	N								534

Significance and number of participants for 10 years old children for the Rod and Frame Test (RFT), Cardio Respiratory Fitness (CRF), Physical Activity (PA), Socio-Economic Status (SES) and NAPLAN data from summed academic scores and individual results in Reading, Writing, and Numeracy. \*Correlation is significant at the 0.05 level (2-tailed). \*\*Correlation is significant at the 0.001 level (2-tailed).

ranged from 130 to 140 participants, whilst field independent participants also ranged from 130 to 140 participants; depending on the variable of interest. Significant differences between the first (field independent) and fourth (field dependent) quartiles obtained from independent-groups *t*-tests for the CRAFT were observed in Numeracy [ $d = 0.82$ ; ( $t_{265}$ ) = 6.71,  $p < 0.001$ ], Overall academic scores [ $d = 0.65$ ; ( $t_{263}$ ) = 5.29,  $p < 0.001$ ], Reading [ $d = 0.47$ ; ( $t_{264}$ ) = 3.859,  $p < 0.001$ ] and Writing [ $d = 0.36$ ; ( $t_{264}$ ) = 2.960,  $p = 0.003$ ] with the field independent

group outperforming the field dependent group on each variable. PA was positively associated with field independent perceptual style [ $d = 0.4$ ; ( $t_{277}$ ) = 3.342,  $p = 0.001$ ] and CRF [ $d = 0.26$ ; ( $t_{215}$ ) = 1.913,  $p = 0.057$ ] There was no significant difference between field independent and field dependent individuals in SES.

Fitter children measured by CRF were associated with enhanced Overall academic scores [ $d = 0.24$ ; ( $t_{262}$ ) = −1.988,  $p = 0.05$ ], Numeracy [ $d = 0.29$ ; ( $t_{266}$ ) = −2.337,  $p = 0.02$ ]

**TABLE 4 |** Hierarchical regression models predicting academic performance in primary schoolchildren.

Model	B	Std. Error	$\beta$	t	Sig.	Adj R <sup>2</sup>	R <sup>2</sup> Change	Sig F Change
<b>Overall academic scores (N = 341)</b>								
(Constant)	476.458	26.841		17.751	0.001			
Step 1 <sub>PA</sub>	0.114	0.304	0.024	0.375	0.708			
CRF	2.253	1.960	0.073	1.150	0.251			
						0.002	0.008	0.275
(Constant)	417.747	27.626		15.122	0.001			
Step 2 <sub>PA</sub>	0.026	0.299	0.005	0.085	0.932			
CRF	1.711	1.895	0.055	0.903	0.367			
SES	24.376	4.263	0.297	5.718	0.001			
GENDER	6.481	6.547	0.055	0.99	0.323			
						0.086	0.089	0.001
(Constant)	457.460	27.214		16.810	0.001			
Step 3 <sub>PA</sub>	0.000	0.285	0.001	0.001	1.000			
CRF	1.253	1.810	0.04	0.692	0.489			
SES	21.381	4.099	0.26	5.216	0.001			
GENDER	0.64	6.325	0.005	0.101	0.92			
CRAFT	-3.335	0.57	-0.297	-5.851	0.001			
						0.168	0.084	0.001
<b>Numeracy (N = 345)</b>								
(Constant)	419.180	32.537		12.883	0.001			
Step 1 <sub>PA</sub>	0.693	0.369	0.118	1.879	0.061			
CRF	0.822	2.382	0.022	0.345	0.73			
						0.011	0.017	0.054
(Constant)	349.968	33.091		10.576	0.001			
Step 2 <sub>PA</sub>	0.3	0.358	0.051	0.839	0.402			
CRF	-0.953	2.27	-0.025	-0.42	0.675			
SES	25.939	5.111	0.257	5.075	0.001			
GENDER (M)	36.167	7.853	0.250	4.606	0.001			
						0.121	0.114	0.001
(Constant)	395.386	32.548		12.148	0.001			
Step 3 <sub>PA</sub>	0.29	0.341	0.049	0.85	0.396			
CRF	-1.538	2.170	-0.04	-0.709	0.479			
SES	22.183	4.923	0.22	4.506	0.001			
GENDER (M)	29.544	7.584	0.205	3.895	0.001			
CRAFT	-3.953	0.679	-0.288	-5.821	0.001			
						0.199	0.079	0.001
<b>Reading (N = 345)</b>								
(Constant)	532.417	33.362		15.959	0.001			
Step 1 <sub>PA</sub>	-0.31	0.378	-0.052	-0.82	0.413			
CRF	1.674	2.425	0.043	0.69	0.49			
						-0.004	0.002	0.68
(Constant)	463.278	34.585		13.396	0.001			
Step 2 <sub>PA</sub>	-0.318	0.373	-0.053	-0.853	0.394			
CRF	1.359	2.358	0.035	0.577	0.565			
SES	28.854	5.295	0.284	5.449	0.001			
GENDER	-0.335	8.153	-0.002	-0.041	0.967			
						0.062	0.059	0.001
(Constant)	503.095	34.69		14.503	0.001			
Step 3 <sub>PA</sub>	-0.343	0.362	-0.057	-0.947	0.344			
CRF	0.87	2.293	0.023	0.379	0.705			
SES	25.844	5.185	0.254	4.984	0.001			
GENDER	-6.11	8.019	-0.042	-0.762	0.447			
CRAFT	-3.349	0.726	-0.239	-4.611	0.001			
						0.101	0.041	0.001

(Continued)

TABLE 4 | Continued

Model	B	Std. Error	$\beta$	t	Sig.	Adj R <sup>2</sup>	R <sup>2</sup> Change	Sig F Change
<b>Writing (N = 346)</b>								
(Constant)	477.854	29.519		16.188	0.001			
Step 1 <sub>PA</sub>	−0.051	0.335	−0.01	−0.153	0.879			
CRF	4.203	2.149	0.122	1.956	0.051			
						0.008	0.014	0.091
(Constant)	434.516	30.994		14.019	0.001			
Step 2 <sub>PA</sub>	0.107	0.334	0.02	0.319	0.75			
CRF	4.642	2.112	0.135	2.198	0.029			
SES	19.219	4.752	0.211	4.044	0.001			
GENDER (F)	−15.500	7.326	−0.119	−2.116	0.035			
						0.072	0.08	0.001
(Constant)	465.396	31.316		14.862	0.001			
STEP 3 <sub>PA</sub>	0.09	0.327	0.017	0.276	0.783			
CRF	4.251	2.070	0.124	2.054	0.041			
SES	16.865	4.69	0.185	3.596	0.001			
GENDER (F)	−20.030	7.261	−0.153	−2.759	0.006			
CRAFT	−2.618	0.657	−0.209	−3.982	0.001			
						0.124	0.054	0.001

Cardio-Respiratory Fitness (CRF), Physical Activity (PA), Percent of Body Fat (% BF), Socio-Economic Status (SES) and Computerised Rod And Frame Test (CRAFT). Bonferroni correction for multiple comparisons set significance at  $p < 0.01$ .

TABLE 5 | High and low competency quartile groups in rod and frame test, cardio-respiratory fitness and physical activity.

Independent Variable	Mean	SD	first Quartile	fourth Quartile	Independent t-tests				
					d	t	df	sig.	Mean Diff
Computer Rod And Frame Test (Subjective visual Vertical)									
Overall academic scores	490	(62.8)	508 (64.3)	468 (58.3)	0.65	5.294	263	0.001	39.6
Numeracy	485	(71.5)	508 (69)	452 (66.8)	0.82	6.711	265	0.001	55.8
Reading	499.5	(77.9)	517 (83)	480 (74.8)	0.47	3.859	264	0.001	37
Writing	483.5	(71.7)	497 (80)	470 (62.9)	0.36	2.960	264	0.003	26.1
Socio-economic status	2.49	(0.72)	2.56 (0.66)	2.42 (0.79)	0.19	1.425	219	0.156	0.14
Cardio-respiratory fitness	5.38	(1.9)	6.2 (2.1)	5.6 (2)	0.26	1.913	215	0.057	0.55
Physical activity	100.1	(11.1)	97.4 (12)	93 (10.8)	0.4	3.342	277	0.001	4.57
Cardio-respiratory fitness									
Overall academic scores	490	(62.8)	498 (60.5)	481 (68.2)	0.24	−1.988	262	0.048	15.8
Numeracy	485	(71.5)	494 (71.2)	473 (77.5)	0.29	−2.337	266	0.02	21.2
Reading	499.5	(77.9)	502 (78.3)	501 (73.7)	0.04	−0.376	267	0.707	3.6
Writing	483.5	(71.7)	495 (65.8)	472 (75.1)	0.33	−2.713	269	0.007	23.2
Physical activity	100.1	(11.1)	93 (10.7)	109 (11.3)	1.4	−12.230	283	0.001	−15.9
Subjective visual vertical	5.31	(4.65)	4.9 (5.1)	6.4 (6.5)	0.2	1.511	234	0.132	1.2
Physical activity									
Overall academic scores	490	(62.8)	496 (56.6)	480 (63)	0.28	−2.519	318	0.012	16.9
Numeracy	485	(71.5)	501 (66.7)	469 (68.8)	0.48	−4.306	322	0.001	32.4
Reading	500	(77.9)	499 (69.3)	495 (89)	0.05	−0.491	320	0.624	4.1
Writing	484	(71.7)	486 (61.9)	473 (80)	0.18	−1.619	324	0.106	12.8
Cardio-respiratory fitness	5.38	(1.9)	6.6 (1.9)	4.3 (1.5)	1.3	−10.920	267	0.001	−2.28
Subjective visual vertical	5.31	(4.65)	5.3 (5.9)	6.2 (5.4)	0.16	1.313	261	0.19	0.92

Polar quartile groupings for Rod and Frame Test (RFT) denoting field independent and field dependent sub-groups with independent variables: Overall academic Scores, Numeracy, Reading, Writing, socio-economic status, physical activity and physical fitness. Cardio-respiratory Fitness and Physical Activity were also separated into quartiles based on high and low Physical Activity and Cardio-respiratory fitness with independent variables: Overall academic Scores, Numeracy, Reading and Writing. The Cohen's d effect sizes for the CRAFT, CRF and PA quartile splits comparisons on Overall Academic Scores, Numeracy, Reading and Writing results. Cohen's d defines small effect as 0.2, medium as 0.5, and large as 0.8.

and Writing [ $d = 0.33$ ; ( $t_{269}$ ) =  $-2.71$ ,  $p = 0.007$ ]. More active children measured using pedometers were associated with Overall academic scores [ $d = 0.28$ ; ( $t_{318}$ ) =  $-2.519$ ,  $p = 0.012$ ] and Numeracy [ $d = 0.48$ ; ( $t_{322}$ ) =  $-4.306$ ,  $p < 0.001$ ].

## DISCUSSION

In this study we examined the view that performance in the rod and frame test is primarily associated with PA measures (Liu and Chepyator-Thomson, 2008, 2009) and less associated with cognitive function as it has been claimed previously that the rod and frame test directly taps into body orientation mechanisms (Liu and Chepyator-Thomson, 2009). If the standard response in predicting the direction of gravitational vertical is for engagement of vestibular and proprioceptive mechanisms no matter the visual angle presented, then various PA variables should have strong and significant associations with the CRAFT measure.

The primary finding here indicates that the ability to accurately estimate gravitational vertical using this CRAFT methodology displayed significant relationships with academic performance, in particular numeracy ability, supporting predictions by Quaiser-Pohl et al. (2004). After controlling for PA and CRF, SES and gender, hierarchical regression analysis provided evidence of significant associations between all measures of academic success and SVV, most strikingly with numeracy. The strength of the relationship between academic success and accuracy in predicting vertical with the current methodology is emphasized by comparison to the effect of SES. CRAFT beta values were all larger than those for SES, except for reading. The association between SVV results and numeracy was the strongest of all the dependent variables measured, explaining 7.9% of the variance of the model. Numeracy had a less strong association, though significant, with SES and male gender. Being female was significantly associated with higher writing test scores.

The proposal that field independent children would have significantly better academic performance compared to that of field dependent children was also supported (Figure 2). All academic measures showed significant positive results for the field independent group, with moderate to large effect sizes in reading ( $d = 0.47$ ) and overall academic scores ( $d = 0.65$ ); and a small to moderate effect size in the writing task ( $d = 0.36$ ). The most significant academic relationship associated with classification of field independent was numeracy, with a large effect size ( $d = 0.82$ ), again providing possible evidence of an association with STEM cluster of cognitive abilities.

An important finding was the relatively small association between PA and CRF with CRAFT scores, where the level of association in both correlation and quartile comparisons was smaller than that found in previous studies. There were no significant correlations between PA and CRF with PV (all  $r > 0.1$ ). In contrast, Liu and Chepyator-Thomson (2008) found Pearson correlations between PA levels and rod and frame test scores (using a 28 degree visual angle) of between  $-0.26$  and  $-0.29$  ( $p < 0.01$ ) in 129 adolescents. In the current

study, when children were separated into field dependent and field independent groups, a significant small to moderate effect size was produced from CRF, PA and percent body fat with the CRAFT. However, the effect size in the current study is significantly lower than what was found in other studies into PA levels using the field dependent – field independent classification (Liu and Chepyator-Thomson, 2009). In the Liu and Chepyator-Thomson (2009) study (using a portable rod and frame test with a larger visual angle of 28 degrees) these researchers found field independent adolescents to be significantly more physically active than field dependent participants and calculated the effect size to be large. The mean for field independent individuals' PA levels was about twice that of FD individuals. In comparison, the present study field dependent - field independent classification provides evidence of a small to moderate positive effect size with PA, with the field independent group scoring 97.4 on the PA index and with field dependent participants averaging 93 in PA index, equating to a 3.5% difference in PA levels. The above findings may relate to the use of a small visual angle CRAFT. As in previous studies, SES was found to have small to moderate correlations with all academic variables (Sirin, 2005).

A possible explanation for the strong association observed here between general academic results and accuracy in predicting vertical, compared with relatively lower association for PA variables, is the measurement method used. The previous positive relationships noted between the “Man in the Frame” rod and frame test and academic variables (Canavan, 1969; Kagan and Zahn, 1975; Kagan et al., 1977) used a small visual angle of 22 degrees and the present study used a small visual angle of 20 degrees. Conversely, studies finding no relationship between the rod and frame test and academic performance have examined PV using the larger visual angle of 28 degrees (Buriel, 1978; Allan et al., 1982; Wong, 1982; Tinajero and Páramo, 1997). The relatively small visual angle employed in the present study means that the test can be classified as a small-scale environment assessment, producing limited visual-vestibular interactions.

Ebenholtz and Benzsawel (1977) first suggested retinal angle in the rod and frame test mediates the results of performance of the rod and frame effect. They further suggest that two related but separate mechanisms are engaged when using small and large visual angle rod and frame test tasks, in their “Dual Process Theory” (Ebenholtz and Glaser, 1982; Coren and Hoy, 1986). Ebenholtz (1977) first suggested that small and large visual angle performances in the rod and frame test are associated with the dual visual projection systems. On the one hand, a small visual angle is related to search fixation, identification, and grasping associated with use of the foveal visual system, whereas a large visual angle rod and frame test is more associated with peripheral visual systems designed for spatial orientation of self and objects in space and for navigation purposes. The Ebenholtz and Glaser (1982) study provides confirmation that large and small frame effects (related to small and large visual angles) are functionally dissimilar and associated with alternate but linked dorsal and ventral visual stream neural processing. Ebenholtz and Glaser (1982) suggest that the small visual angle RFT engages a foveal



stimulated ventral visual stream sensory input, whilst a large visual angle produces a dorsal visual stream peripheral vision stimulus.

Streibel and Ebenholtz (1982) compared the performance of the rod and frame test using small and large visual angles and found a significant stimulus size (visual angle) effect, with the small visual angle rod and frame test (9 degrees) having a larger mean score than the large angle rod and frame test (41 degrees). Further, the Streibel and Ebenholtz (1982) findings support previous reports of poor correlations between perceptions arising from small and large scale environments (Spinelli et al., 1999; Quaiser-Pohl et al., 2004). In addition to this Quaiser-Pohl et al. (2004) concluded that a distinction should be made between children's spatial cognition in small-scale and large-scale spatial environments, and that they measure definably different elements of spatial ability.

Supporting neurophysiological evidence is provided by a study by Lopez et al. (2011) using a rod and frame test with a small visual angle of 14 degrees, examining the sequence of neurologic activation patterns from electro-encephalograph event potentials during the rod and frame task. They found early activation in the right temporo-occipital cortex ventral stream; likely activating visual-visual mechanisms in the extrastriate cortex engaged at around 75 milliseconds and implicating attentional processes. Later, combined ventral stream and dorsal stream visual-vestibular zones are activated at around 260 milliseconds bilaterally in the temporal, occipital, and parietal areas. These results suggest engagement of temporal cortical zones are implicated in the frame effect for the purpose of maintaining internal estimates of vertical, while the parietal cortical areas are more involved with the control of posture, actions, and visuospatial processing.

We propose that the PV in the current study was most likely reliant on participants' previous experiences of vertical alignment from vestibular and somato-sensory stimulation, then stored as an internal representation in temporal and parietal cortical zones. This requires significant cortical involvement to accurately predict gravitational vertical (Lopez et al., 2011). It is further proposed that children most capable of cognitive transformation of vestibular and somato-sensory inputs from previous experience into a stored internal representation of gravitational vertical have greater school academic success, particularly in numeracy. The observation of neural ventral visual stream engagement in a small angle rod and frame test (Lopez et al., 2011) may help explain the lack relationship found in the current study between CRAFT results with PA, CRF and percent of body mass measures. The small retinal angle produces cognitive responses more attuned to small scale environments and related cognitive processes (Quaiser-Pohl et al., 2004). It is possible enriched environmental experiences engaging small visual angle tasks may have developmental influences in forming this internal representation, together with inherited genetic advantage.

A second possible pathway modulating the association between CRAFT methodology examining SVV and academic performance is via attentional abilities, as outlined in previous studies (Blowers and O'Connor, 1978; Shinar et al., 1978; Berger

and Goldberger, 1979; Agathos et al., 2015). The attentional theory may complement the proposition of small visual angle perception, in that the small field environment is proposed to stimulate attentional mechanisms more effectively (Quaiser-Pohl et al., 2004).

In this study, children's performance on a test to determine visual vertical using a CRAFT provided evidence of a moderate to strong association with numeracy. Also, a field independent perceptual style was associated with a strong effect size in numeracy, compared to field dependent categorized children. These findings of relationships between subjective vertical perception and numeracy may be explained in a number of ways. First, the neural origins of the internal construct of an internal representation of vertical alignment arise within cortical zones located in the temporal and parietal cortex (Lopez et al., 2011), an area shared with the neural foundations for analysis of dimensional magnitudes and the "approximate number system"; thought to provide the foundation of numeracy skills (Walsh, 2003). Secondly, Witkin et al. (1977) suggest that field independent mode in the perception of upright is associated with competence in restructuring the spatial field and associated with an articulated body concept. Factor analytical studies have also found PV loads significantly with activities associated with spatial restructuring (Witkin and Goodenough, 1981). Further, evidence supports the association between spatial restructuring and numeracy skills (Evans et al., 2013).

The moderate to strong relationships found in correlation and hierarchical regression, as well as between classification as having a field independent perceptual style and numeracy, suggest a link with STEM strands of academic tasks. The current study confirms previous findings, suggesting that children who accurately estimate vertical alignment are likely to perform at a higher level and may subsequently specialize in STEM fields (Witkin et al., 1975, 1977; Kent-Davis and Cochran, 1989; Evans et al., 2013).

However, being field independent is also an important advantage in reading, writing and for combined academic ability (overall academic scores), thus supporting the view that the current CRAFT methodology taps into cognitive processes involved in elements of general intelligence, as indicated by previous literature (Goodenough and Karp, 1961; Blowers and O'Connor, 1978; Berger and Goldberger, 1979; Agathos et al., 2015). In addition, the current findings add weight to the Dubois and Cohen (1970) proposal that the rod and frame test provides an analysis of a characteristic of intelligence "not contaminated by complex spatial reasoning," and also support the view that general academic performance is predicted by the field independent classification, with no academic advantage in being field dependent.

Because eleven year old children develop from being relatively field dependent to relatively more field independent, it is possible that this transition period is particularly susceptible to a small visual angle rod and frame test. Dubois and Cohen (1970) conducted a study on 143 female undergraduate university students and found field independent participants out-performed field dependent participants and had a proclivity for STEM

subjects. Unfortunately, the visual angle used in the above study was not provided, making it difficult to draw further conclusions.

Correlation analysis provided evidence of a significant but small association between CRF and PA with a number of academic variables. However, in the present study, elevated levels of CRF and PA were significantly associated with higher overall academic scores and writing score. Enhanced numeracy scores associated with higher levels of PA ( $d = 0.48$ ) and CRF ( $d = 0.29$ ) were the most significant; with no significant association found for reading, writing or overall academic scores. Support for the finding of an association between CRF solely with numeracy comes from previous research (Chaddock-Heyman et al., 2015; Huang et al., 2015). It appears that substantially increased levels of CRF are required before academic results in numeracy are positively influenced, a finding supported by previous research (Sibley and Etnier, 2003; Etnier et al., 2006; Drollette et al., 2014; Chaddock-Heyman et al., 2015).

The association between spatial processing ability of PV with numerical cognition may provide support for theories of embodied cognition (Koziol et al., 2012) which may well have evolved as an exaptive trait in the evolution of arboreal pre-humans, in which upright orientation was formulated from highly unpredictable compliant body support (on tree branches) in a visually complex and ambiguous environment made up of geometric mechanical shapes and forms; unstable and unreliable as a sensory frame of reference – unlike our modern terrestrial based world (Haynes et al., 2017). This theory builds on the proposals firstly by Thorpe et al. (2007) suggesting the arboreal origins of human uprightness; and by Godfrey-Smith (2002) arguing there is an association between environmental complexity and cognition; and proposals by Povinelli and Cant (1995) who contend arboreal requirements of large-bodied primates require complex sensorimotor adaptations that have led to self-awareness and other cognitive traits. Further, the “Dual Process Theory” (Ebenholtz and Glaser, 1982; Coren and Hoy, 1986) Ebenholtz (1977) may help explain the evolutionary construction differentiating small and large visual angle spatial perceptions. Engaging ventral (foveal) visual streams with small visual angle SVV possibly evolved in pre-human primates for upright search, fixation and identification of food sources embedded within the complex geometric arboreal field, with hand alignments to hold branches to support upright orientation. Large visual angle SVV tasks may have evolved to exploit peripheral (dorsal stream) visual systems designed for spatial orientation framed along the gravitational axis when the pre-human primate navigated and negotiated complex three dimensional spatial geometric environments on non-direct pathways containing gaps and obstacles. Pre-human primates may have then used many non-upright postures with spatial memory abilities to return to the nesting colony.

## Limitations

Necessarily, the cross-sectional design of the present study does not allow causal relations to be determined between SVV and STEM performance. Further, the pedometer measurements

employed here are broad measures of PA and do not measure activity style or intensity. Some PA data collection technologies are more sensitive to the intensity load and style of PA, but these were not available to this study, and this constitutes a limitation. Finally, it is possible the instructions given to children in the conduct of the rod and frame test may have affected the study outcomes. However, the use of the rocket ship analogy and the time spent in describing and practicing the test before measurements were taken were consistent features throughout.

## CONCLUSION

The current study provides evidence that error scores from a small visual angle CRAFT are more strongly related to academic test results than are PA and CRF variables. Associations were found between the accuracy in predicting gravitational vertical and academic performance, particularly numeracy. Current results support the view that the existing methodology provides an important indicator of STEM potential. In terms of style of spatial processing, field independent children, compared to field dependent children, had better academic results with numeracy, exhibiting a strong effect size and emphasizing FI as a gauge of STEM potential. Field independent participants also had increased PA and CRF levels (small to moderate effect size). Finally, enhanced levels of CRF and PA (based on quartile splits) compared to lower levels of CRF and PA showed a significant small to moderate effect size in performance on the overall academic scores, numeracy and writing tests.

Whilst this is a cross-sectional analysis and no causal relationships can be attributed, the strength of the results provides grounds to suggest that possible PA pathways in intervention activities focusing on small scale visuo-perceptual tasks could influence academic performance. Supporting this view, recent evidence suggests that both motor activities (Adams et al., 2014) and musical training (Norton et al., 2005) can improve spatial cognition. A recent meta-review concluded that motor performance in combat sports, musical instrument studies and gymnastics (but not ballet) had the greatest impact on spatial abilities (Voyer and Jansen, 2017). Further, Newcombe (2017) proposes that spatial ability is malleable and suitable training may enhance STEM abilities. Future studies could therefore examine physical activities in pre-pubescent children who were engaging in spatial perception in both small and large field environments (small and large visual angles) to stimulate perceptual switching abilities, navigation and cognitive processing.

The strength of the findings here raises questions about the types of PA interventions that may influence vertical perception ability, and consequently academic results, and are scope for future research. For example, play environments with off-the-ground beam-like stepping bridges with rope hand supports forming complex geometric shapes to hold and orientate toward (stimulating small and large visual angles) are worthy of further

investigation. Older children may benefit from single path map reading orienteering in a complex natural environmental setting combining both large and small field environments spatial perceptions with moderate to high levels of PA intensity. Further, elements of physical activities designed to stimulate alignment and orientation to vertical should logically include short periods of PA requiring moderate to high levels of intensity – for example soccer or tennis training drills engaging in small field visuo-spatial tasks, such as ball-at-feet soccer drills (Pesce, 2012) and yoga poses with visual focus and attention directed onto body parts (Diamond and Ling, 2016).

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Australian Capital Territory Health and Community Care Human Research Ethics Committee. The protocol was approved by the Australian Capital Territory Health and Community Care Human Research Ethics Committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## REFERENCES

- Adams, D. M., Stull, A. T., and Hegarty, M. (2014). Effects of mental and manual rotation training on mental and manual rotation performance. *Spat. Cognit. Comput.* 14, 169–198. doi: 10.1080/13875868.2014.913050
- Agathos, C. P., Bernardin, D., Huchet, D., Scherlen, A. C., Assaiante, C., and Isableu, B. (2015). Sensorimotor and cognitive factors associated with the age-related increase of visual field dependence: a cross-sectional study. *Age* 37:9805. doi: 10.1007/s11357-015-9805-x
- Allan, R., Roberts, E., and McIntire, G. (1982). The influence of intensive training in decoding skills on field dependence-independence in first grade children. *Res. Rural Educ.* 1.
- Amador-Campos, J. A., and Kirchner-Nebot, T. (1999). Correlations among scores on measures of field dependence-independence cognitive style, cognitive ability, and sustained attention. *Percept. Mot. Skills* 88, 236–239. doi: 10.2466/pms.1999.88.1.236
- Arbuthnot, J. (1972). Cautionary note on measurement of field independence. *Percept. Mot. Skills* 35, 479–488. doi: 10.2466/pms.1972.35.2.479
- Assaiante, C., and Amblard, B. (1993). Ontogenesis of head stabilization in space during locomotion in children: influence of visual cues. *Exp. Brain Res.* 93, 499–515. doi: 10.1007/BF00229365
- Bagust, J., Docherty, S., Haynes, W., Telford, R., and Isableu, B. (2013). Changes in rod and frame test scores recorded in schoolchildren during development—a longitudinal study. *PLoS One* 8:e65321. doi: 10.1371/journal.pone.0065321
- Barra, J., Marquer, A., Joassin, R., Reymond, C., Metge, L., Chauvineau, V., et al. (2010). Humans use internal models to construct and update a sense of verticality. *Brain* 133, 3552–3563. doi: 10.1093/brain/awq311
- Beets, M. W., Patton, M. M., and Edwards, S. (2005). The accuracy of pedometer steps and time during walking in children. *Med. Sci. Sports Exerc.* 37, 513–520. doi: 10.1249/01.mss.0000155395.49960.31
- Berger, E., and Goldberger, L. (1979). Field dependence and short-term memory. *Percept. Mot. Skills* 49, 87–96. doi: 10.2466/pms.1979.49.1.87
- Blowers, G. H., and O'Connor, K. P. (1978). Relation of eye movements to errors on the rod-and-frame test. *Percept. Motor Skills* 46, 719–725. doi: 10.2466/pms.1978.46.3.719
- Brady, F. (1995). Sports skill classification, gender, and perceptual style. *Percept. Mot. Skills* 81, 611–620. doi: 10.2466/pms.1995.81.2.611

## HUMAN PARTICIPANTS APPROVAL STATEMENT

This study was approved by the Australian Capital Territory Health and Community Care Human Research Ethics Committee.

## AUTHOR CONTRIBUTIONS

WH was the lead author, designed the research, carried out the data analysis, and worked on the conceptual design. GW was involved in the authorship, data analysis, and statistical modeling. RA contributed to the statistical design, data analysis, and paper construction. BI contributed to the research design and paper construction.

## ACKNOWLEDGMENTS

We thank the teachers, parents, and children for their willingness to participate in this project, and the Commonwealth Education Trust for their financial contribution.

- Buriel, R. (1978). Relationship of three field-dependence measures to the reading and math achievement of Anglo American and Mexican American children. *J. Educ. Psychol.* 70, 167. doi: 10.1037/0022-0663.70.2.167
- Canavan, D. (1969). *Field Dependence in Children As a Function of Grade, Sex, and Ethnic Group Membership*. Washington, DC: American Psychological Association.
- Carver, S., Kiemel, T., and Jeka, J. J. (2006). Modeling the dynamics of sensory reweighting. *Biol. Cybern.* 95, 123–134. doi: 10.1007/s00422-006-0069-5
- Chaddock-Heyman, L., Erickson, K. I., Kienzler, C., King, M., Pontifex, M. B., Raine, L. B., et al. (2015). The role of aerobic fitness in cortical thickness and mathematics achievement in preadolescent children. *PLoS One* 10:e0134115. doi: 10.1371/journal.pone.0134115
- Clemens, I. A., De Vrijer, M., Selen, L. P., Van Gisbergen, J. A., and Medendorp, W. P. (2011). Multisensory processing in spatial orientation: an inverse probabilistic approach. *J. Neurosci.* 31, 5365–5377. doi: 10.1523/JNEUROSCI.6472-10.2011
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Coren, S., and Hoy, V. S. (1986). An orientation illusion analog to the rod and frame: relational effects in the magnitude of the distortion. *Percept. Psychophys.* 39, 159–163. doi: 10.3758/BF03212486
- Davis, J. K., and Cochran, K. F. (1982). *Toward an Information Processing Analysis of Field-Independence*. Washington, D.C: ERIC.
- Diamond, A., and Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Dev. Cognit. Neurosci.* 18, 34–48. doi: 10.1016/j.dcn.2015.11.005
- Dreyer, A. S., Dreyer, C. A., and Nebelkopf, E. B. (1971). Portable rod-and-frame test as a measure of cognitive style in kindergarten children. *Percept. Mot. Skills* 33, 775–781. doi: 10.2466/pms.1971.33.3.775
- Drollette, E. S., Scudder, M. R., Raine, L. B., Moore, R. D., Saliba, B. J., Pontifex, M. B., et al. (2014). Acute exercise facilitates brain function and cognition in children who need it most: an ERP study of individual differences in inhibitory control capacity. *Dev. Cognit. Neurosci.* 7, 53–64. doi: 10.1016/j.dcn.2013.11.001
- Dubois, T. E., and Cohen, W. (1970). Relationship between measures of psychological differentiation and intellectual ability. *Percept. Mot. Skills* 31, 411–416. doi: 10.2466/pms.1970.31.2.411



- Duncan, S. C., Duncan, T. E., Strycker, L. A., and Chaumeton, N. R. (2002). Neighborhood physical activity opportunity: a multilevel contextual model. *Res. Quart. Exerc. Sport* 73, 457–463. doi: 10.1080/02701367.2002.10609046
- Ebenholtz, S. M. (1977). Determinants of the rod and frame effect: the role of retinal size. *Percept. Psychophys.* 22, 531–538. doi: 10.3758/bf03198759
- Ebenholtz, S. M., and Benzschawel, T. L. (1977). The rod and frame effect and induced head tilt as a function of observation distance. *Percept. Psychophys.* 22, 491–496. doi: 10.3758/BF03199516
- Ebenholtz, S. M., and Glaser, G. W. (1982). Absence of depth processing in the large-frame rod-and-frame effect. *Percept. Psychophys.* 32, 134–140. doi: 10.3758/BF03204272
- Etnier, J. L., Nowell, P. M., Landers, D. M., and Sibley, B. A. (2006). A meta-regression to examine the relationship between aerobic fitness and cognitive performance. *Brain Res. Rev.* 52, 119–130. doi: 10.1016/j.brainresrev.2006.01.002
- Evans, C., Richardson, J. T., and Waring, M. (2013). Field independence: reviewing the evidence. *Br. J. Educ. Psychol.* 83, 210–224. doi: 10.1111/bjep.12015
- Ganley, C. M., Vasilyeva, M., and Dulaney, A. (2014). Spatial ability mediates the gender difference in middle school students' science performance. *Child Dev.* 85, 1419–1432. doi: 10.1111/cdev.12230
- Godfrey-Smith, P. (2002). "Environmental complexity and the evolution of cognition," in *The Evolution of Intelligence*, eds R. J. Sternberg and J. C. Kaufman (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 223–249.
- Golomer, E., Crémieux, J., Dupui, P., Isableu, B., and Ohlmann, T. (1999). Visual contribution to self-induced body sway frequencies and visual perception of male professional dancers. *Neurosci. Lett.* 267, 189–192. doi: 10.1016/S0304-3940(99)00356-0
- Goodenough, D. R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychol. Bull.* 83, 675–694. doi: 10.1037/0033-2909.83.4.675
- Goodenough, D. R., and Karp, S. A. (1961). Field dependence and intellectual functioning. *J. Abnor. Soc. Psychol.* 63, 241–246. doi: 10.1037/h0046524
- Gough, H. G., and Olton, R. M. (1972). Field independence as related to nonverbal measures of perceptual performance and cognitive ability. *J. Consult. Clin. Psychol.* 38:338.
- Haynes, W., Waddington, G., and Adams, R. (2017). Physical activity, academic and developmental measures in older primary school-children: a principal components analysis. *Cogent. Psychol.* 4:1413928. doi: 10.1080/23311908.2017.1413928
- Haynes, W., Bagust, J., and McGrath, M. (2008). "Perception of vertical improves over time in young children as measured by a computer rod and frame test (CRAFT)," in *Proceedings of the Commonwealth Games Sports Medicine Conference*, Melbourne, VIC.
- Hodgson, C. I., Christian, E., and McMorris, T. (2010). Performance on the portable rod and frame test predicts variation in learning the kayak roll. *Percept. Mot. Skills* 110, 479–487. doi: 10.2466/pms.110.2.479-487
- Huang, T., Tarp, J., Domazet, S. L., Thorsen, A. K., Froberg, K., Andersen, L. B., et al. (2015). Associations of adiposity and aerobic fitness with executive function and math performance in Danish adolescents. *J. Pediatr.* 167, 810–815. doi: 10.1016/j.jpeds.2015.07.009
- Isableu, B., Ohlmann, T., Crémieux, J., and Amblard, B. (1997). Selection of spatial frame of reference and postural control variability. *Exp. Brain Res.* 114, 584–589. doi: 10.1007/pl00005667
- Isableu, B., Ohlmann, T., Crémieux, J., Vuillerme, N., Amblard, B., and Gresty, M. A. (2010). Individual differences in the ability to identify, select and use appropriate frames of reference for perceptuo-motor control. *Neuroscience* 169, 1199–1215. doi: 10.1016/j.neuroscience.2010.05.072
- Jia, S., Zhang, Q., and Li, S. (2014). Field dependence-independence modulates the efficiency of filtering out irrelevant information in a visual working memory task. *Neuroscience* 278, 136–143. doi: 10.1016/j.neuroscience.2014.07.075
- Kagan, S., and Zahn, G. L. (1975). Field dependence and the school achievement gap between anglo-American and Mexican-American children. *J. Educ. Psychol.* 67, 643–650. doi: 10.1037/0022-0663.67.5.643
- Kagan, S., Zahn, G. L., and Gealy, J. (1977). Competition and school achievement among Anglo-American and Mexican-American children. *J. Educ. Psychol.* 69, 432–441. doi: 10.1037/0022-0663.69.4.432
- Kent-Davis, J., and Cochran, K. F. (1989). An information processing view of field dependence-independence. *Early Child Dev. Care* 51, 31–47. doi: 10.1080/0300443890510104
- Kozioł, L. F., Budding, D. E., and Chidekel, D. (2012). From movement to thought: executive function, embodied cognition, and the cerebellum. *Cerebellum* 11, 505–525. doi: 10.1007/s12311-011-0321-y
- Liu, W., and Chepyator-Thomson, J. R. (2008). Associations among field dependence-independence, sports participation, and PA level among school children. *J. Sport Behav.* 31:130.
- Liu, W., and Chepyator-Thomson, J. R. (2009). Field dependence-independence and PA engagement among middle school students. *Phys. Educ. Sport Pedag.* 14, 125–136. doi: 10.1080/17408980801974960
- Lopez, C., Mercier, M. R., Halje, P., and Blanke, O. (2011). Spatiotemporal dynamics of visual vertical judgments: early and late brain mechanisms as revealed by high-density electrical neuroimaging. *Neuroscience* 181, 134–149. doi: 10.1016/j.neuroscience.2011.02.009
- Maceachron, A. E., and Gruenfeld, L. W. (1978). The effects of family authority structure and socioeconomic status on field independence. *J. Soc. Psychol.* 104, 49–56. doi: 10.1080/00224545.1978.9924037
- MacLeod, C. M., Jackson, R. A., and Palmer, J. (1986). On the relation between spatial ability and field dependence. *Intelligence* 10, 141–151. doi: 10.1016/0160-2896(86)90011-5
- MacNeilage, P. R., Banks, M. S., Berger, D. R., and Bulthoff, H. H. (2007). A Bayesian model of the disambiguation of gravito-inertial force by visual cues. *Exp. Brain Res.* 179, 263–290. doi: 10.1007/s00221-006-0792-0
- Medendorp, W. P., and Selen, L. J. (2017). Vestibular contributions to high-level sensorimotor functions. *Neuropsychologia* 105, 144–152. doi: 10.1016/j.neuropsychologia.2017.02.004
- Meek, F., and Skubic, V. (1971). Spatial perception of highly skilled and poorly skilled females. *Percept. Mot. Skills* 33, 1309–1310. doi: 10.2466/pms.1971.33.3f.1309
- Mergner, T., Maurer, C., and Peterka, R. J. (2003). A multisensory posture control model of human upright stance. *Prog. Brain Res.* 142, 189–201. doi: 10.1016/S0079-6123(03)42014-1
- Mezzacappa, E. (2004). Alerting, orienting, and executive attention: developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Dev.* 75, 1373–1386. doi: 10.1111/j.1467-8624.2004.00746.x
- NAPLAN (2010). *National Assessment Program [Homepage on the Internet]*. Canberra: Australian Curriculum Assessment and Reporting Authority. National Assessment Program: Literacy and Numeracy. Available at: <https://www.nap.edu.au/> [accessed November 27, 2010].
- Newcombe, N. (2017). "Harnessing Spatial Thinking to Support Stem Learning" OECD Education Working Papers, No. 161, Paris: OECD. Available at: <http://dx.doi.org/10.1787/7d5dcae6-en>
- Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., and Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain Cognit.* 59, 124–134. doi: 10.1016/j.bandc.2005.05.009
- Oltman, P. K. (1968). A portable rod-and-frame apparatus. *Percept. Motor Skills* 26, 503–506. doi: 10.2466/pms.1968.26.2.503
- Pesce, C. (2012). Shifting the focus from quantitative to qualitative exercise characteristics in exercise and cognition research. *J. Sport Exerc. Psychol.* 34, 766–786. doi: 10.1123/jsep.34.6.766
- Poitras, V. J., Gray, C. E., Borghese, M. M., Carson, V., Chaput, J. P., Janssen, I., et al. (2016). Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. *Appl. Physiol. Nutr. Metab.* 41, S197–S239. doi: 10.1139/apnm-2015-0663
- Povinelli, D. J., and Cant, J. G. (1995). Arboreal clambering and the evolution of self-conception. *Quart. Rev. Biol.* 70, 393–421. doi: 10.1086/419170
- Quaiser-Pohl, C., Lehmann, W., and Eid, M. (2004). The relationship between spatial abilities and representations of large-scale space in children—a structural equation modeling analysis. *Personal. Indi. Diff.* 36, 95–107. doi: 10.1016/s0191-8869(03)00071-0
- Raviv, S., and Nabel, N. (1990). Relationship between two different measurements of field-dependence and athletic performance of adolescents. *Percept. Mot. Skills* 70, 75–81. doi: 10.2466/pms.1990.70.1.75

- Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* 6, 15–32. doi: 10.1214/ss/1177011926
- Ruiz-Ariza, A., Grao-Cruces, A., de Loureiro, N. E. M., and Martínez-López, E. J. (2017). Influence of physical fitness on cognitive and academic performance in adolescents: a systematic review from 2005–2015. *Int. Rev. Sport Exerc. Psychol.* 10, 108–133. doi: 10.1080/1750984X.2016.1184699
- Schuler, J. R., Bockisch, C. J., Straumann, D., and Tarnutzer, A. A. (2010). Precision and accuracy of the subjective haptic vertical in the roll plane. *BMC Neurosci.* 11:83. doi: 10.1186/1471-2202-11-83
- Shea, D. L., Lubinski, D., and Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. *J. Educ. Psychol.* 93:604. doi: 10.1037/0022-0663.93.3.604
- Shinar, D., McDowell, E. D., Rackoff, N. J., and Rockwell, T. H. (1978). Field dependence and driver visual search behavior. *Hum. Factors* 20, 553–559. doi: 10.1177/001872087802000505
- Sibley, B. A., and Etnier, J. L. (2003). The relationship between physical activity and cognition in children: a meta-analysis. *Pediatr. Exerc. Sci.* 15, 243–256. doi: 10.1123/pes.15.3.243
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev. Educ. Res.* 75, 417–453. doi: 10.3102/00346543075003417
- Smith, J. J., Eather, N., Morgan, P. J., Plotnikoff, R. C., Faigenbaum, A. D., and Lubans, D. R. (2014). The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis. *Sports Med.* 44, 1209–1223. doi: 10.1007/s40279-014-0196-4
- Spinelli, D., Antonucci, G., Daini, R., Martelli, M. L., and Zoccolotti, P. (1999). Hierarchical organisation in perception of orientation. *Perception* 28, 965–979. doi: 10.1068/P280965
- Streibel, M. J., and Ebenholtz, S. M. (1982). Construct validity of perceptual style: role of stimulus size in the embedded-figures test and the rod-and-frame test. *Percept. Psychophys.* 31, 128–138. doi: 10.3758/BF03206212
- Tarnutzer, A. A., Bockisch, C., Straumann, D., and Olasagasti, I. (2009). Gravity dependence of subjective visual vertical variability. *J. Neurophysiol.* 102, 1657–1671. doi: 10.1152/jn.00007.2008
- Telford, R. D., Bass, S. L., Budge, M. M., Byrne, D. G., Carlson, J. S., Coles, D., et al. (2009). The lifestyle of our kids (LOOK) project: outline of methods. *J. Sci. Med. Sport* 12, 156–163. doi: 10.1016/j.jsams.2007.03.009
- Thorpe, S. K., Holder, R. L., and Crompton, R. H. (2007). Origin of human bipedalism as an adaptation for locomotion on flexible branches. *Science* 316, 1328–1331. doi: 10.1126/science.1140799
- Tinajero, C., and Páramo, M. F. (1997). Field dependence-independence and academic achievement: a re-examination of their relationship. *Br. J. Educ. Psychol.* 67, 199–212. doi: 10.1111/j.2044-8279.1997.tb01237.x
- Tomkinson, G. R., Léger, L. A., Olds, T. S., and Cazorla, G. (2003). Secular trends in the performance of children and adolescents (1980–2000). *Sports Med.* 33, 285–300. doi: 10.2165/00007256-200333040-00003
- Tomprowski, P. D., McCullick, B., Pendleton, D. M., and Pesce, C. (2015). Exercise and children's cognition: the role of exercise characteristics and a place for metacognition. *J. Sport Health Science* 4, 47–55. doi: 10.1016/j.jshs.2014.09.003
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., and Newcombe, N. S. (2017). I. Spatial skills, their development, and their links to mathematics. *Monogr. Soc. Res. Child Dev.* 82, 7–30. doi: 10.1111/mono.12280
- Voyer, D., and Jansen, P. (2017). Motor expertise and performance in spatial tasks: a meta-analysis. *Hum. Mov. Sci.* 54, 110–124. doi: 10.1016/j.humov.2017.04.004
- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cognit. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002
- Wang, L., Cohen, A. S., and Carr, M. (2014). Spatial ability at two scales of representation: a meta-analysis. *Learn. Ind. Diff.* 36, 140–144. doi: 10.1016/j.lindif.2014.10.006
- Witkin, H. A. (1959). The perception of the upright. *Sci. Am.* 200, 50–70. doi: 10.1038/scientificamerican0259-50
- Witkin, H. A., and Asch, S. E. (1948). Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. *J. Exp. Psychol.* 38, 762–782. doi: 10.1037/h0053671
- Witkin, H. A., and Goodenough, D. R. (1981). *Cognitive Styles, Essence and Origins: Field Dependence and Field Independence Psychological Issues*. New York, NY: International University Press.
- Witkin, H. A., Goodenough, D. R., and Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. *J. Pers. Soc. Psychol.* 7, 291–300. doi: 10.1037/h0025070
- Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. (1975). Field-dependent and field-independent cognitive styles and their educational implications. *ETS Res. Bull. Ser.* 1975, 1–64. doi: 10.1002/j.2333-8504.1975.tb01065.x
- Witkin, H. A., Moore, C. A., Oltman, P. K., Goodenough, D. R., Friedman, F., Owen, D. R., et al. (1977). Role of the field-dependent and field-independent cognitive styles in academic evolution: a longitudinal study. *J. Educ. Psychol.* 69, 197–211. doi: 10.1037/0022-0663.69.3.197
- Wong, P. C. N. (1982). *Relationship of Two Field-dependent-independent Measures to Reading and Math Achievement of Anglo-American, First-generation Chinese-and Mexican-American Elementary School Children* (Dissertation), PC Wong.

**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 Haynes, Waddington, Adams and Isableu. 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.



# Numerical and Non-numerical Predictors of First Graders' Number-Line Estimation Ability

Richard J. Daker and Ian M. Lyons\*

Department of Psychology, Georgetown University, Washington, DC, United States

## OPEN ACCESS

### Edited by:

Sharlene D. Newman,  
Indiana University Bloomington,  
United States

### Reviewed by:

Koen Luwel,  
KU Leuven, Belgium  
Luis J. Fuentes,  
Universidad de Murcia, Spain

### \*Correspondence:

Ian M. Lyons  
ian.lyons@georgetown.edu

### Specialty section:

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

**Received:** 16 August 2018

**Accepted:** 07 November 2018

**Published:** 30 November 2018

### Citation:

Daker RJ and Lyons IM (2018)  
Numerical and Non-numerical  
Predictors of First Graders'  
Number-Line Estimation Ability.  
Front. Psychol. 9:2336.  
doi: 10.3389/fpsyg.2018.02336

Children's ability to map numbers into a spatial context has been shown to be a powerful predictor of math performance. Here, we investigate how three types of cognitive abilities – approximate number processing ability, symbolic number processing ability, and non-numerical cognitive abilities – predict 0–100 number-line estimation performance in first graders. While each type of measure predicts number-line performance when considered individually, when considered together, only symbolic number comparison and non-verbal reasoning predicted unique variance in number-line estimation. Moreover, the relation between symbolic number comparison and number-line ability was stronger for male students than for female students, suggesting potential gender differences in the way boys and girls accomplish mapping numbers into space. These results suggest that number-line estimation ability is largely reflective of the precision with which symbolic magnitudes are represented (at least among boys). Our findings therefore suggest that promoting children's understanding of symbolic, rather than non-symbolic, numerical magnitudes may help children learn better from number-lines in the classroom.

**Keywords:** number-line estimation, spatial processing, early numeracy, gender differences, number symbols

## INTRODUCTION

Children's ability to map numbers into a spatial context has been shown to be a powerful predictor of math performance (Siegler and Booth, 2004; Booth and Siegler, 2008; Sasanguie et al., 2013; Lyons et al., 2014; Friso-van den Bos et al., 2015; Schneider et al., 2018). Past research using number-line estimation tasks, in which children mark the spatial location of a given number (e.g., "72") on a horizontal line (typically with only the endpoints indicated, e.g., with 0 at the left end and 100 at the right end), has been shown to predict performance on other measures of basic numeracy (Laski and Siegler, 2007; Maertens et al., 2016) and arithmetic (Siegler and Booth, 2004; Booth and Siegler, 2008; Lyons et al., 2014; Schneider et al., 2018). Moreover, experimental research has demonstrated that playing board games meant to bolster the visuospatial representation of numerical values in children improves numerical knowledge and performance on a range of numerical and mathematical tasks (Ramani and Siegler, 2008; Siegler and Ramani, 2009; Ramani et al., 2012; Maertens et al., 2016). The precision with which children perform number-line estimation tasks has been argued to reflect the precision with which children represent numerical magnitudes (Laski and Siegler, 2007; Booth and Siegler, 2008), which has been proposed by some researchers to serve as a key foundation for more complex mathematical processing (e.g., Feigenson et al., 2013; Siegler and Braithwaite, 2017). Given both the predictive and potentially causal role

that visuospatial representations of numerical magnitude play in the development of mathematics, an important question is what basic numerical abilities contribute to the early development of these visuospatial representations.

Past work examining number-line estimation ability has in part focused on pinpointing when key developmental shifts occur (Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2008; Siegler and Ramani, 2009). Of particular importance, multiple studies have found that by the time children are in second grade, students have developed a fairly linear 0–100 mental number-line, whereas children in first grade are, on average, still in the process of linearizing their visuospatial representations of 0–100 (Siegler and Booth, 2004; Booth and Siegler, 2006). The development of this mental number-line has been theorized to have a core role in broader numerical cognition (Siegler et al., 2011, 2013). Siegler et al. (2011) have argued for an integrated theory of numerical development in which numerical development involves coming to understand that “all real numbers have magnitudes that can be ordered and assigned specific locations on number-lines.”

While the development of a precise mental number-line is thought to play an important role in broader numerical development, it is important to note that performance on the number-line task is not a pure reflection of children’s numerical understanding. Recent work has shown that non-numerical factors, particularly strategy selection, play a substantial role in children’s number-line performance (Barth and Paladino, 2011; Cohen and Blanc-Goldhammer, 2011; Slusser et al., 2013; Rouder and Geary, 2014; Dackermann et al., 2015; Peeters et al., 2016, 2017; van’t Noordende et al., 2016). The role that individual differences in strategy selection play in number-line performance makes it important to consider non-numerical factors, such as non-verbal reasoning ability, that may impact children’s performance on the number-line task.

Given the centrality with which increasing precision of the mental number-line is theorized to play in more general numerical development, understanding what basic numerical and non-numerical cognitive abilities predict the ability to precisely map numbers into space during key developmental shifts can give us insight into possible mechanisms that could underlie core numerical abilities. The goal of the present research is to understand what basic numerical and non-numerical cognitive abilities predict the ability to precisely map numbers into space during a key developmental period.

Here we consider three main hypotheses about what types of basic numerical and non-numerical cognition may support visuospatial number-line estimates in early grade school. According to one view, approximate number processing has been argued to be the foundation upon which more complex numerical abilities are grounded (Dehaene, 1997; Libertus et al., 2011, 2012, 2013; Feigenson et al., 2013). Because number-line estimation abilities are still developing in first graders (Siegler and Booth, 2004; Booth and Siegler, 2006), it may be the case that individual differences in approximate number processing at this age are predictive of number-line abilities. More specifically, this view predicts that a common measure of approximate number processing (i.e., determining which of two arrays of dots contains

the greater quantity) should be a robust predictor of number-line estimation accuracy.

A second view is that symbolic representation of numerical quantities (e.g., Indo-Arabic numerals) serves as a crucial conceptual leap that underpins much of the subsequent development of more complex numerical thinking (e.g., De Smedt et al., 2009; Bugden and Ansari, 2011; Merkley and Ansari, 2016; Vanbinst et al., 2016; Núñez, 2017). A canonical measure of basic symbolic number processing is via numeral comparison tasks in which children indicate which of two numerals (e.g., ‘6’ and ‘8’) represents the greater quantity. Performance on this task has been shown to be a strong predictor of math achievement across a wide range of ages and settings (Holloway and Ansari, 2009; Nosworthy et al., 2013; Vanbinst et al., 2016; Sasanguie et al., 2017; Lyons et al., 2018). Moreover, previous work has shown that improvements in number-line estimation accuracy are associated with improvements in numeral comparison ability (Laski and Siegler, 2007; Ramani and Siegler, 2008), indicating that these two basic numerical abilities may be fundamentally intertwined early in development. However, it remains less clear whether these two abilities are *uniquely* related – that is, does the relation obtain even after controlling, for example, for approximate number processing, general cognitive ability, and other basic numerical abilities such as counting, ordering and estimation.

A third hypothesis is that reasoning or general cognitive ability – more so than other basic numerical abilities – is the strongest predictor of number-line estimation in early grade-school. As the work demonstrating effects of strategy utilization shows (e.g., Slusser et al., 2013; Peeters et al., 2016), numerical understanding is not the only thing that contributes to number-line performance. It is therefore possible that children with higher levels of general reasoning ability will demonstrate better number-line performance (even after controlling for basic numerical abilities), via the ability to select the most effective strategies.

Of course, the hypotheses outlined above are not mutually exclusive. Indeed, previous studies have demonstrated that measures of all three kinds significantly relate to number-line performance (Opfer and Siegler, 2007; Sasanguie et al., 2012; Fuhs and McNeil, 2013; Fazio et al., 2014; Maertens et al., 2016). However, to our knowledge, no work has examined the *unique* contributions of these numerical and non-numerical abilities to number-line estimation. Learning what predicts unique variance in number-line estimation ability will allow for a more precise understanding of which aspects of early numeracy are foundational in the development of a precise mental number-line. Such an understanding would allow for the generation of testable hypotheses about how to improve number-line estimation ability (and in turn math skills).

While the measures mentioned above are of primary theoretical interest, assessing the extent to which other basic numerical abilities (i.e., numerical ordering ability or counting proficiency) predicts number-line estimation performance comes with at least two benefits: First, it is possible that the three hypotheses outlined above are incomplete – testing other basic abilities allows us to check for additional factors that may



impact number-line estimation performance not covered by those hypotheses. Second, given that other basic numerical abilities have also been shown to predict more complex math (Lyons and Beilock, 2011; Lyons et al., 2014), it is important to control for these other abilities to estimate as precisely and conservatively as possible the unique variance in number-line performance that can be attributed to the measures of primary theoretical interest outlined above.

In this study, we used data from over 200 Dutch first graders to understand what basic numerical and general cognitive factors predict unique variance in 0–100 number-line performance. We chose to focus on first graders because past work has suggested that important developmental shifts in 0–100 number-line performance occur during this year (Siegler and Booth, 2004; Booth and Siegler, 2006), and because this age group shows sufficient variability in terms of individual differences in our sample to allow for meaningful inferences to be drawn from a multiple regression approach. Finally, given substantial evidence for gender differences in number-line estimation, especially in first grade (Thompson and Opfer, 2008; Gunderson et al., 2012; Hutchison et al., 2018), we assess whether the strength of the potential relations between basic numerical and non-numerical cognitive abilities and number-line estimation depends on (i.e., interacts with) gender.

## MATERIALS AND METHODS

### Participants

235 Dutch children (105 female; mean age = 7.06 years; *SD* age = 0.44) in first grade participated. Of this initial sample, 24 were removed from analysis for chance performance on any of the tasks and another 3 were removed for scores on any task that were greater than 4 standard deviations away from the mean. Of the initial sample of 235, 27 were removed (11.5%) for a total analytic sample size of 208 (97 female).

It is important to note that the data reported here are part of a larger data set, some of which has been reported on in previous work (e.g., Lyons et al., 2014). Crucially, both the theoretical questions addressed and the analyses described here are novel.

### Procedure

The ethics review board at Maastricht University approved the data collection procedure used in this study. Children came from seven different primary schools in the Netherlands, where data collection took place. The schools provided written notification of the purpose and nature of the data collection procedures to parents. Parents could withhold consent by returning the appropriate form. All data were collected one-on-one by trained project workers at the children's schools. All data were collected in one session. All measures were computerized with the exception of the non-verbal intelligence measure (Ravens), which was in a paper-and-pencil format. Before each numerical task,

participants were given 3–6 practice trials. During the main experimental trials, no feedback was given for any of the tasks.

### Primary Tasks of Interest

#### Number-Line Estimation (NumLine)

In the NumLine task, children were shown a horizontal line with 0 marked on the left side and 100 marked on the right. On each trial, participants saw an Arabic numeral centered above the line and heard the same number over headphones. Their task was to click where on the number-line the target number should be placed based on the quantity it represented. All stimuli remained on the screen until the child responded. Children completed a total of 26 trials. Reliability on this task was high:  $\alpha = 0.90$ .

Consistent with previous research on the 0–100 number-line task (Siegler and Booth, 2004; Booth and Siegler, 2006), performance on this task was near ceiling for children above first grade in the broader dataset from which this study is drawn. Ceiling-level performance dramatically reduces variability of scores in older children, making individual-differences-based results with this task in older children largely uninterpretable. On the other hand, we did see substantial variability in performance among first graders; this coupled with the observation that meaningful developmental changes are still occurring on this task in first graders (see Introduction) prompted us to focus on first graders for the purposes of the present research.

#### Numerical Comparison (NumComp)

In the NumComp task, children were shown two Arabic numerals presented horizontally, and their task was to decide which number was greater. A total of 64 trials were presented, comprised of 32 one-digit and 32 two-digit trials. Four ratio ( $R = \min/\max$ ) ranges were used:  $R \leq 0.5$ ,  $R = 0.5$ ,  $0.5 < R < 0.7$ , and  $R \geq 0.7$ . Each ratio range occurred equally across one- and two-digit trials. All stimuli remained on the screen until the child responded. Reliability on this task was high:  $\alpha = 0.92$ .

#### Dot Comparison (DotComp)

In the DotComp task, children were shown two dot arrays, and their task was to decide which array contained more dots. 64 trials were presented, and quantities and ratios used were identical to those in the NumComp task. Overall area and average individual dot-size were always incongruent with number such that the array with fewer dots always had greater overall area and larger average dot-size. This was done to preclude participants from using strategies based on surface area or dot size to determine which array contained the greater quantity of dots. Additional stimulus details for this task, including manipulation checks, can be found in Lyons et al. (2014). All stimuli remained on the screen until the child responded. Reliability on this task was high:  $\alpha = 0.92$ .

#### Non-verbal Intelligence (Ravens)

The Ravens task is a normed, timed, visuospatial reasoning test for children (Raven et al., 1995). A colored pattern appeared and children were asked to select the missing piece out of six choices. The task was comprised of a total of 36 trials, and the total

number answered correctly was the child's score. Van Bon (1986) reported reliabilities of 0.80 or higher for the Dutch version of this task.

## Additional Numerical Tasks of Secondary Interest and Covariates

### Numerical Ordering (NumOrd)

In the NumOrd task, children were shown three single-digit Arabic numerals presented horizontally. On half of the trials, the three numbers were in increasing order from left to right. On the other half of trials, numbers were either in decreased or mixed order. Children were instructed to indicate with a button press whether the numbers were in increasing order or not. All stimuli remained on the screen until the child responded. The 28 trials were roughly divided into distances of 1–3. For example, an in-order trial with distance 1 may contain the numbers “4, 5, and 6” whereas an in-order trial with distance 3 may contain the numbers “2, 5, and 8.” Reliability on this task was high:  $\alpha = 0.82$ .

### Object Matching (ObjMatch)

In the ObjMatch task, children were presented with a sample array of common objects (including animals and fruits) and two test arrays. The children's task was to select the test array that contained the same number of items as the sample array. A total of 45 trials were shown: in 15 trials, all objects in each of the arrays were the same; in 15 trials, each array contained different types of objects (but the objects within an array were of the same type); and in the remaining 15 trials, each array contained a mixture of object types. The number of objects in the arrays ranged from 1 to 6, and the difference in the number of objects between the two test arrays was 1 or 2. All stimuli remained on the screen until the child responded. Reliability on this task was high:  $\alpha = 0.92$ .

### Dot Quantity Estimation (DotEst)

In the DotEst task, children saw a single array of dots presented for a very short time (750 ms) – too quickly to be counted individually – followed by a visual mask. The task was to estimate the amount of dots present in the array with a verbal response, which was manually recorded by the experimenter. This task contained a total of 84 trials, made up of 12 trials each with the quantities 1, 2, 3, 4, 7, 11, and 16. Note that results do not substantially change if only quantities 7, 11, and 16 are used. Reliability on this task was acceptable:  $\alpha = 0.76$ .

### Counting (Counting)

In the Counting task, children were presented with between 1 and 9 dots, and their task was to count the number of dots as quickly and accurately as possible. This task contained a total of 45 trials, 5 with each quantity. Children responded verbally, and their responses were manually recorded by the experimenter. Children were instructed to press a button as they gave their response in order to estimate response times. Reliability on this task was high:  $\alpha = 0.90$ .

### Visual-Audio Matching (VisAud)

In the VisAud task, children heard a number word spoken aloud and were immediately presented with an Arabic number on the screen. The task was to indicate by button press whether the numbers were the same. This task was comprised of 64 trials, half involving one-digit numbers and the other half involving two-digit numbers. On trials in which the numbers did not match, the ratio between the numbers ranged from 0.25 to 0.89. Moreover, non-matching trial stimuli avoided tens-ones confusion items (e.g., 32 and “twenty-three”). Reliability on this task was high:  $\alpha = 0.90$ .

### Reading Ability (Reading)

The Reading task was part of the Maastricht Dyslexia Differential Diagnosis battery (Blomert and Vaessen, 2009). Children completed three subtasks that contained high-frequency words, low-frequency words, or pseudo-words. For each subtask, participants were shown up to five screens, each with up to 15 items, for a total of 75 items per subtask. Children were tasked with reading each item aloud as quickly and accurately as possible in 30 s. This task was included to control for basic reading fluency in the multiple regression analyses. The Reading score was the total number of words correctly read across each subtask. Test-retest reliability reported for this task is 0.95 (Blomert and Vaessen, 2009).

### Basic Stimulus-Response Processing (StimResp)

In the StimResp task, children were presented with four boxes arranged horizontally on the screen. On each trial, a fish appeared in one of the four boxes, and the children's task was to press the corresponding key on the response box as quickly and accurately as they could. Children completed a total of 20 trials. This task was included to control for basic stimulus-response processing in the multiple regression analyses. All stimuli remained on the screen until the child responded. Reliability on this task was high:  $\alpha = 0.88$ .

### Task Scoring

For the NumLine and the DotEst task, we used percent absolute errors:  $PAE = |Est - Target| / Scale$ , where *Est* is the child's estimate, *Target* is the target number, and *Scale* is the range of target numbers. The range was 100 for the NumLine task and 16 for DotEst. For the NumLine task, note that results were highly similar if degree of linearity (a child's  $R^2$  indicating the linear fit between their estimates and the actual value) was used instead of PAE. A higher value thus indicates *poorer* performance on these tasks; for this reason, values were multiplied by  $-1$  before being entered into regression models.

For tasks in which error rate and response time data was available (NumComp, DotComp, Counting, NumOrd, VisAud, ObjMatch, and StimResp), we used a composite of error rates and response times on correct trials:  $P = RT(1 + 2ER)$ , where *RT* is a child's mean response-time for that task and *ER* is the child's error-rate for that task (Lyons et al., 2014). This was done to account for speed-accuracy tradeoffs and to cut down on the number of analyses required, thus minimizing

the risk of Type 1 errors. A higher value thus indicates *poorer* performance on these tasks; for this reason, values were multiplied by  $-1$  before being entered into regression models.

We used total number of correct responses for both the Ravens and Reading tasks, hence a higher value indicates *better* performance on these tasks.

## RESULTS

### Basic Descriptives

**Table 1** shows mean performance levels for each task (before multiplying relevant scores by  $-1$ ), and **Figure 1** shows zero-order correlations between all measures (and Age).

### Unique Predictors of Number-Line Estimation

We first entered all numerical measures, all non-numerical measures, and a dummy variable for gender (0 = male, 1 = female) into a regression model to predict NumLine performance. Age was also included as a control measure. **Table 2** shows results of the initial model, and **Figure 2** visualizes relative partial correlation coefficients taken from the multiple-regression model. Results of the initial model show that only NumComp, Ravens, Gender, and Age explain unique variance in NumLine performance.

We next aimed to identify the most parsimonious model possible by removing predictors that failed to predict unique variance in NumLine performance, removing predictors with the lowest  $p$ -values in a step-wise fashion until all predictors were significant at  $p < 0.05$ . **Table 3** shows the progression of model reduction. In the process of model reduction, all predictors were removed with the exception of NumComp, Ravens, and Gender. Because of significant theoretical interest

in the DotComp task, we decided to retain it in the final model (shown in **Table 4**) despite its not predicting unique variance in NumLine performance (indeed, it would have been the second predictor omitted in the process of model reduction). Age was also retained as an important control variable despite not predicting unique variance in NumLine performance.

### Modulation by Gender

In this section, we assessed whether the relations between the predictors of interest retained in the final model (NumComp, DotComp, and Ravens) and NumLine were modulated by (interacted with) gender. To do so, we ran a model predicting NumLine in which we interacted NumComp, DotComp, and Ravens with gender. Results (shown in **Table 5**) demonstrate a significant NumComp  $\times$  Gender interaction ( $p = 0.023$ ). Results did not show a significant interaction with gender for either DotComp or Ravens (both  $ps > 0.45$ ).

To decompose the significant NumComp  $\times$  Gender interaction, we next ran multiple-regression models predicting NumLine from NumComp, DotComp, and Ravens, separately by gender. Results (plotted in **Figure 3** and shown in **Table 6**) show that while NumComp was the strongest predictor of NumLine for boys, it did not predict unique NumLine variance for girls. Note that Ravens was a significant predictor for both boys and girls; DotComp was not significant for either.

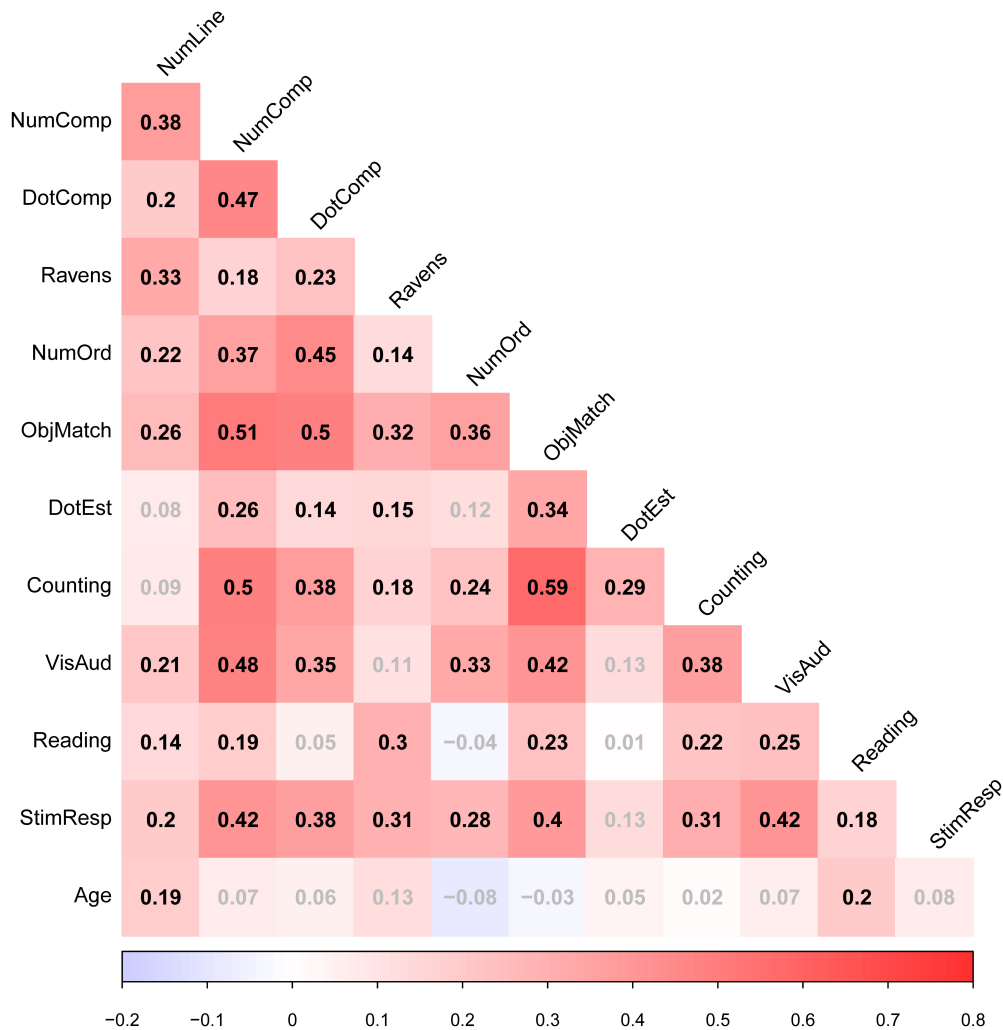
## DISCUSSION

Across a range of ages and contexts, children's ability to map numbers into a spatial context has been shown to be a powerful predictor of math performance (Siegler and Booth, 2004; Booth and Siegler, 2008; Sasanguie et al., 2013; Lyons et al., 2014; Friso-van den Bos et al., 2015; Schneider et al., 2018). The goal of the present work was to assess which numerical and non-numerical cognitive abilities predict unique variance in 0–100 number-line estimation ability in first graders. Results indicated that symbolic number processing, but not non-symbolic number processing, predicted unique variance in number-line estimation ability. Moreover, within the realm of symbolic number processing, it was numerical magnitude comparison that was predictive of unique number-line variance, while other symbolic measures, like numeral ordering, did not predict unique variance. The number-line task has been conceptualized as indexing children's underlying representation of numerical magnitude (Laski and Siegler, 2007; Booth and Siegler, 2008); the present work suggests that number-line estimation is indeed best predicted by measures of numerical magnitude. Crucially, however, our work here indicates that this interpretation is specific to measures of *symbolic* magnitude representation. Furthermore, results showed that non-verbal reasoning ability also predicted unique variance in number-line estimation, suggesting a role for non-numeric, domain-general cognitive abilities in number-line performance. Interestingly, we also found that the relationship between number-line estimation ability and numeral comparison ability was modulated by gender such that numeral comparison was predictive for boys, but not girls. Our results help clarify the

**TABLE 1 |** Descriptives.

Predictor	Mean
<i>N</i>	208 (97female)
NumLine <sup>1</sup>	14.20 (0.39)
NumComp <sup>2</sup>	1738 (29)
DotComp <sup>2</sup>	1721 (31)
Ravens <sup>3</sup>	25.6 (0.3)
NumOrd <sup>2</sup>	4945 (127)
ObjMatch <sup>2</sup>	4979 (93)
DotEst <sup>1</sup>	8.46 (0.20)
Counting <sup>2</sup>	3665 (62)
VisAud <sup>2</sup>	1898 (28)
Reading <sup>3</sup>	49.9 (1.8)
StimResp <sup>2</sup>	970 (12)
Age	7.06 (0.03)

Values in parentheses are SE of the mean unless indicated otherwise. Superscripts refer to scoring metrics: <sup>1</sup>percent absolute error, <sup>2</sup>composite measure of ER and RT, and <sup>3</sup>number correct. See Materials and Methods for further details on these scoring metrics.



**FIGURE 1 |** Zero-order correlation matrix. The figure shows zero-order correlations.

nature of the numerical magnitude representations indexed by number-line estimation tasks in early grade-school. Moreover, as number-lines are a ubiquitous visualization device found in early mathematics classrooms, our results may also point to practical implications for the kinds of basic abilities that permit children to get the most out of this common pedagogical tool.

Recent work has demonstrated that symbolic and non-symbolic representations of quantity are distinct in both adults and young children (Verguts et al., 2005; Carey et al., 2017; Lyons et al., 2012, 2015, 2018). Siegler and colleagues have argued that the number-line task assesses underlying magnitude representations (Laski and Siegler, 2007; Booth and Siegler, 2008), but until this point it hasn't been explicitly tested whether the theorized underlying magnitude is symbolic or non-symbolic. Our findings show it is *symbolic* magnitude comparison that predicts unique variance in number-line estimation ability, whereas approximate magnitude comparison does not. If the number-line task reflected the representational precision of non-symbolic quantities (i.e., the width or narrowness of

non-symbolic tuning curves), the NumLine task should have shown a strong relation with children's ability to distinguish between two non-symbolic magnitudes (indexed here via the DotComp task). However, our results indicated this was not the case. Instead, we found that number-line estimation precision is more closely associated with children's ability to judge the relative magnitudes represented by number *symbols* (indexed here via the NumComp task). Our results thus clarify an important point with respect to a prominent view of what is indexed by number-line estimation tasks (Siegler and Braithwaite, 2017). Namely, while our results are broadly consistent with the view that number-line tasks primarily index relative magnitude processing (Laski and Siegler, 2007; Booth and Siegler, 2008), here we add the important caveat that the operative notion of magnitude is primarily the *symbolic* aspect of numerical magnitude.

An important question that follows is what exactly is meant by symbolic numerical magnitude (at least in the present context)? As noted above, recent work has indicated that the meaning of



**TABLE 2** | Initial multiple regression model.

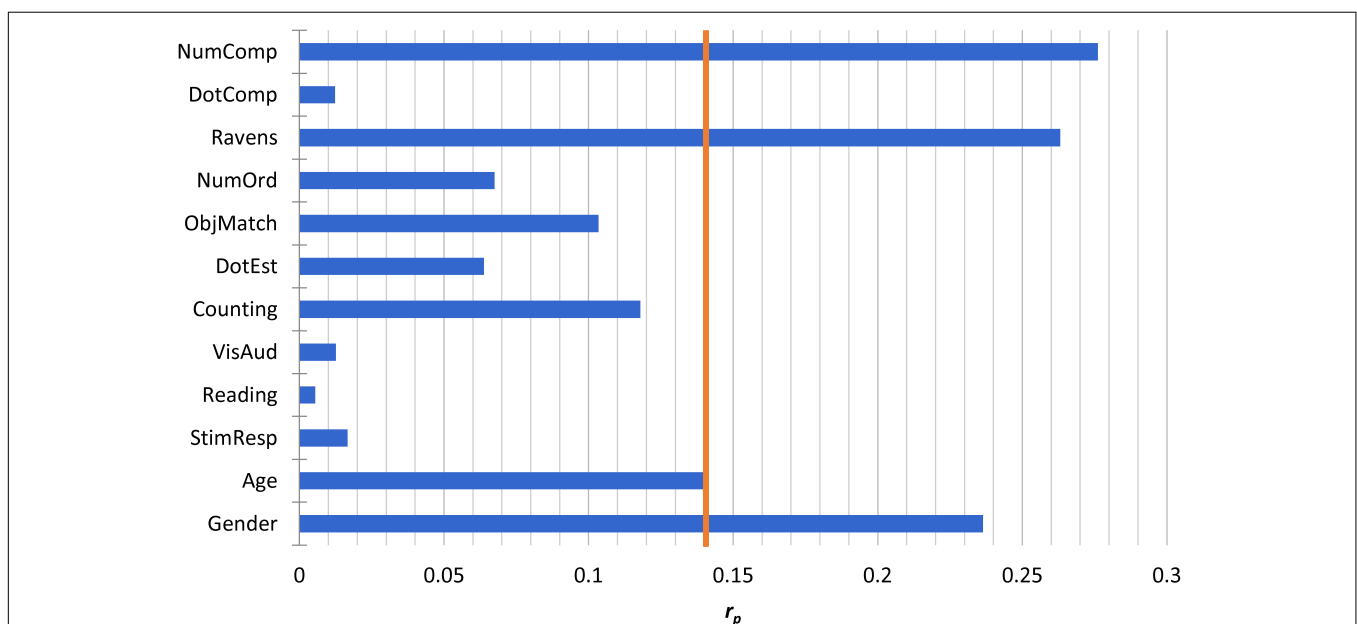
DV: NumLine					
Predictor	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	<i>r<sub>p</sub></i>
NumComp	4.9E – 3	1.1E – 3	4.01	8.6E – 5	0.276
DotComp	–1.6E – 4	9.6E – 4	–0.17	0.864	–0.012
Ravens	0.35	0.09	3.81	1.9E – 4	0.263
NumOrd	2.0E – 4	2.1E – 4	0.94	0.346	0.067
ObjMatch	5.2E – 4	3.6E – 4	1.45	0.148	0.103
DotEst	–11.45	12.83	–0.89	0.373	0.064
Counting	–8.3E – 4	5.0E – 4	–1.66	0.099	–0.118
VisAud	1.8E – 4	1.0E – 3	0.18	0.861	0.013
Reading	–1.1E – 3	0.01	–0.08	0.940	–0.005
StimResp	–5.6E – 4	2.4E – 3	–0.23	0.817	–0.017
Age	1.57	0.79	1.97	0.050	0.140
Gender	–2.45	0.72	–3.40	8.2E – 4	0.236

Overall adjusted  $R^2 = 0.276$ , numerator  $df = 1$  for each predictor. Error (denominator)  $df = 195$ .  $r_p$  = partial- $r$  value.

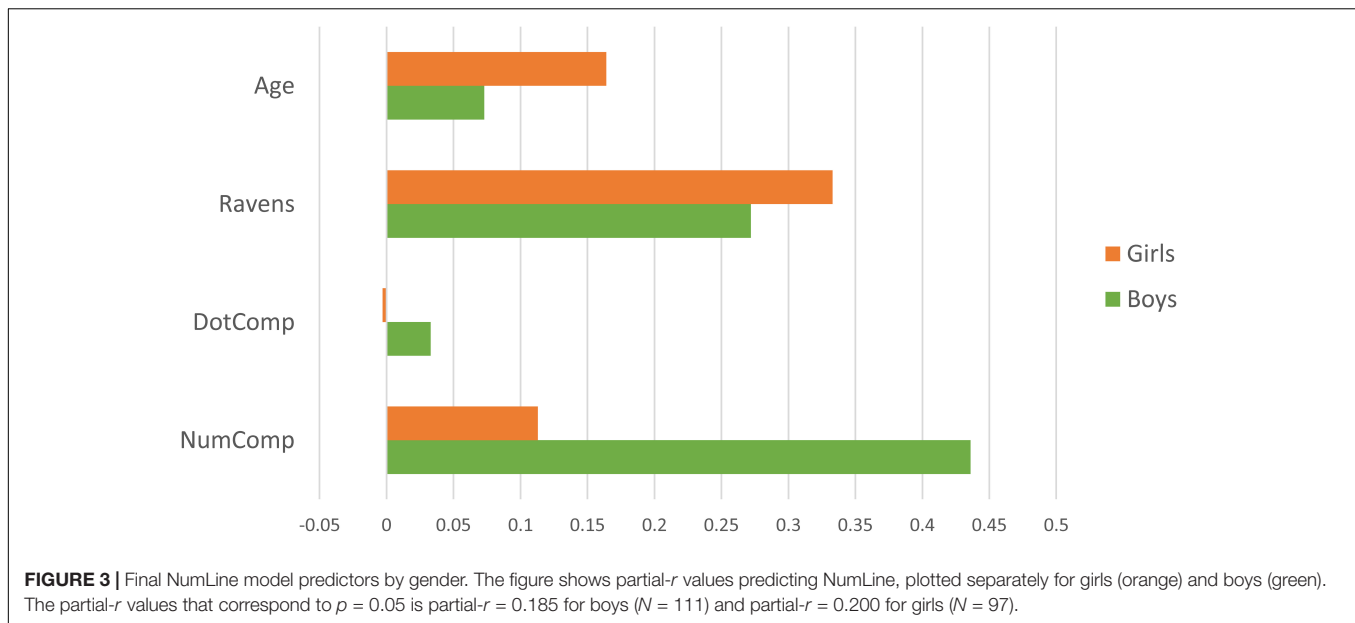
number symbols is likely relatively distinct from approximate magnitudes (Verguts et al., 2005; Lyons et al., 2012, 2015, 2018; Carey et al., 2017), and our results here are broadly consistent with this. In response, some have proposed that number symbols are primarily associative in nature, drawing much of their meaning from associations (such as relative order – ‘What comes next?’) with other number symbols (Nieder, 2009; Núñez, 2017; Lyons and Beilock, 2018). However, in the current context of understanding number-line estimation, this associative aspect of number symbols does not appear to be the critical factor either, as we failed to find that performance on the symbolic number ordering task (NumOrd) predicts unique NumLine variance. An

alternative hypothesis proposed by Verguts et al. (2005); see also Roggeman et al. (2007) is that exact representation of numbers (as is thought to be the case with number symbols) operates via ‘place coding.’ Numbers are represented with equal precision regardless of numerical magnitude and indexed based on their relative position on a putative internal mental number-line. Perhaps, most intriguingly here, this mental number-line is typically conceptualized in an explicitly visuospatial manner. If it were the case that, rather than just serving as a useful metaphor, children may actually represent numerical magnitudes by placing numbers along a mental line. In such a framework, the precision with which a given quantity is placed on this *mental* line should translate directly to the precision with which it is placed on an *external* line, as in number-line tasks. It may be that first graders rely on this place-based coding to represent symbolic quantities. Hence, this place-based coding may underlie both their ability to compare symbolic magnitudes and generate number-line estimates, as indicated by the strong unique relation between these two tasks we see here.

Here we also found that non-verbal reasoning ability (Ravens) predicted unique variance in number-line performance, suggesting a role for non-numerical cognitive ability in number-line estimation. Previous work on number-line estimation has found individual differences in strategy use (Booth and Siegler, 2008; Slusser et al., 2013; Dackermann et al., 2015; Peeters et al., 2016; van’t Noordende et al., 2016), so a potential interpretation of this relation is that stronger non-verbal reasoning skills may allowing children to select more effective strategies. One practical implication is that future work using the number-line estimation task should take care to control for non-verbal reasoning ability in order to ensure that any claims made about the number-line task are not



**FIGURE 2** | Unique NumLine predictors. The figure shows partial- $r$  values from the initial (full) model (see **Table 2**) predicting NumLine performance. The vertical line indicates the partial- $r$  value corresponding to  $p = 0.05$ .

**TABLE 3 |** Progression of model reduction.

Step	Predictor removed	$p$ -value of removed predictor	Adjusted $R^2$ after predictor removed	Change in adjusted $R^2$ from initial model
Initial	—	—	0.27609	—
1	Reading	0.940	0.27976	+0.00367
2	VisAud	0.869	0.28332	+0.00723
3	StimResp	0.842	0.28680	+0.01070
4	DotEst	0.369	0.28747	+0.01138
5	NumOrd	0.325	0.28757	+0.01148
6	ObjMatch	0.149	0.28369	+0.00760
7	Counting	0.193	0.28118	+0.00509

*DotComp would have been the second predictor omitted ( $p = 0.870$  after reading was removed) but was retained due to significant theoretical interest in this task.*

unknowingly driven by its relation with non-verbal reasoning. From a theoretical perspective, the finding that both numerical magnitude representation and non-verbal reasoning ability each predict unique variance in number-line estimation suggests that both types of ability (numerical and non-numerical) work in conjunction to support effective number-line estimation.

Another result of potential interest here is that the relation between symbolic magnitude comparison and number-line estimation ability was modulated by gender: while this relation obtained for boys ( $r_p = 0.436$ ), it did not for girls ( $r_p = 0.164$ ). Given the preceding discussion, one question is thus why girls did not show a significant relation between NumComp and NumLine performance. Boys consistently show a higher spatial skills on average than girls (Voyer et al., 1995; Kimura, 1999; Terlecki and Newcombe, 2005; Feng et al., 2007). Therefore, one possibility is that, owing to lower general spatial skills, girls are less likely on average than boys to develop an explicitly visuospatial place-coding representation of numerical magnitude, or girls may

**TABLE 4 |** Final model details.

DV: NumLine					
Predictor	$b$	se	$t$	$p$	$r_p$
NumComp	4.1E – 3	8.9E – 4	4.60	7.5E – 6	0.308
DotComp	3.7E – 4	8.5E – 4	0.44	0.664	0.031
Ravens	0.36	0.08	4.30	2.6E – 5	0.290
Age	1.27	0.76	1.67	0.096	0.117
Gender	–2.86	0.68	–4.22	3.7E – 5	–0.285

*Overall adjusted  $R^2 = 0.281$ , numerator  $df = 1$  for each predictor. Error (denominator)  $df = 202$ .  $r_p$  = partial- $r$  value.*

do so later in development than boys. Consistent with this notion, previous work has found an advantage for boys in number-line estimation (Hutchison et al., 2018). For boys, the number-line task is already cognitively aligned to the spatial manner in which they represent numbers. By contrast, if girls do not primarily represent numbers spatially, there will be an additional cost of translating from a non-spatial representation in order to plot a number in space. Moreover, this putative difference in number representation would also explain the lack of a unique relation between numerical magnitude representation (as indexed by the NumComp task) and number-line performance among girls. If girls do not represent numerical magnitudes spatially, then the ability that allows them to compare symbolic magnitudes would not relate to the ability to plot numbers on a line.

While the idea that boys and girls may vary in the extent to which their representations of numerical magnitudes are spatial in nature is admittedly a *post hoc* interpretation of our results, it does generate some useful hypotheses that may guide future work. First, it suggests that boys' performance on a number-line task would be harmed more by visuospatial load or by changing the format of the number line (from horizontal

**TABLE 5 |** Gender interaction model.

DV: NumLine					
Predictor	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	<i>r<sub>p</sub></i>
NumComp	5.6E – 3	1.1E – 3	4.97	1.5E – 6	0.332
DotComp	3.4E – 4	1.1E – 3	0.32	0.748	0.022
Ravens	0.34	0.12	2.90	0.004	0.202
Age	1.27	0.76	1.67	0.096	0.112
Gender	–13.01	6.06	–2.15	0.033	–0.150
NumComp * Gender	–4.1E – 3	1.83E – 3	–2.24	0.026	–0.157
DotComp * Gender	–3.5E – 4	0.001	–0.20	0.844	–0.014
Ravens * Gender	0.09	0.17	0.57	0.570	0.040

Overall adjusted  $R^2 = 0.321$ , numerator  $df = 1$  for each predictor. Error (denominator)  $df = 199$ .  $r_p$  = partial- $r$  value.

**TABLE 6 |** Separate models by gender.

DV: NumLine						
	Predictor	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	<i>r<sub>p</sub></i>
Boys	NumComp	5.7E – 3	1.1E – 3	4.99	2.4E – 6	0.436
	DotComp	3.6E – 4	1.1E – 3	0.34	0.735	0.033
	Ravens	0.34	0.116	2.91	0.004	0.272
	Age	0.73	0.98	0.75	0.456	0.073
Girls	NumComp	1.6E – 3	1.4E – 3	1.09	0.279	0.113
	DotComp	–4.8E – 5	1.4E – 3	–0.03	0.974	–0.003
	Ravens	0.41	0.122	3.39	0.001	0.333
	Age	1.92	1.21	1.60	0.113	0.164

Boys: Overall adjusted  $R^2 = 0.302$ , numerator  $df = 1$  for each predictor. Error (denominator)  $df = 106$ . Girls: Overall adjusted  $R^2 = 0.162$ , numerator  $df = 1$  for each predictor. Error (denominator)  $df = 92$ .  $r_p$  = partial- $r$  value.

to vertical, for instance) than girls' performance (controlling for general spatial ability). Second, it may be the case that differences in general spatial ability may explain the gender difference in performance on number-line tasks and the simultaneous absence of gender differences on less explicitly spatial measures of numeracy (Hutchison et al., 2018). Moreover, differences in the extent to which representations of numerical magnitude are spatial may also have an impact on how well children learn about numbers and math from spatial pedagogical strategies (discussed below). Finally, it should be noted that we did not find a gender interaction for the Ravens task, suggesting that non-numerical cognitive abilities – regardless of how symbolic magnitudes are being represented – play a similar role for boys and girls.

In addition to informing theories of number-line estimation and informing debates on broader numerical development, we note that the present work has potential implications for educational settings. Number-lines of course arise not just in the context of the eponymous cognitive task, but they are a common pedagogical tool found in early grade-school classrooms used to promote development of numerical understanding. While experimental work would need to be done to lend greater support to this idea, our work suggests that working to promote children's understanding of *symbolic*, rather than non-symbolic, numerical magnitudes may help children get more out of number-lines as a pedagogical tools. Importantly,

however, this may be qualified by gender, applying more strongly to male than female children, on average. Finally, the finding that non-verbal reasoning ability predicts number-line estimation ability (regardless of gender) also suggests that children with lower non-numerical reasoning skills may require additional support when using number-lines as pedagogical tools.

Finally, it is important to note the limitations of the present study. First, this study deals with just one number-line range (0–100). This was done for the practical reason that the majority of children in this age range are familiar with two-digit numbers, but not all may be comfortable with three-digit numbers, so using the range 0–100 is perhaps best suited for the majority of students at this age. Second, the data reported here focused on a single age-range (first graders). This was because number-line estimation ability on 0–100 tasks is still developing for children of this age. As such, focusing on this age range and task presented an opportunity to investigate factors that may affect the development of number-line estimation ability. Furthermore, it should be noted that previous work on number-line estimation has shown that findings from different age groups and different number-line ranges have generalized well to one another (e.g., Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2008; Siegler and Ramani, 2009). One might argue a third potential limitation is that our findings were biased to show effects of

symbolic number comparison over non-symbolic number comparison because the target magnitudes in the number-line task were presented as symbols rather than dot arrays. However, we controlled for several other symbolic measures, including ordering and number-naming, and, it should be noted, none of those predicted unique variance. This suggests that the effect of numeral comparison we found is not merely driven by the fact that it shares a format with the target magnitude.

## CONCLUSION

Our work shows that unique variance in number-line estimation ability is explained by individual differences in symbolic magnitude processing and non-verbal reasoning ability, but not approximate magnitude processing. This finding refines theories of number-line estimation by clarifying that the representations of numerical magnitude tapped by the number-line task appears to be largely symbolic in nature rather than reflecting the degree of representational precision of approximate tuning curves. However, the relation between performance on a symbolic magnitude task and number-line estimation was found to be stronger for boys than girls, potentially due to differences in the degree to which number representations are spatial in nature among boys and girls. This work suggests that promoting children's understanding of symbolic, rather than non-symbolic, numerical magnitudes may help children learn better from number-lines in the

classroom and that future research should treat number-line estimation tasks as reflecting underlying representations of symbolic magnitude.

## AUTHOR CONTRIBUTIONS

Data for the study came from a preexisting dataset. All authors contributed to the conception of the study and writing of the manuscript. RD completed the data analysis.

## FUNDING

This research was supported by funding from the Canadian Institutes of Health Research (CIHR), The Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Research Chairs Program (CRC) to Daniel Ansari, as well as Departmental Start-Up Funds (Georgetown University) to IL. Data collection costs were paid in part by Boom Test Uitgevers Amsterdam BV.

## ACKNOWLEDGMENTS

Special thanks to Daniel Ansari for providing data access and to Daniel Ansari and Jane Hutchison for helpful comments on an earlier version of the manuscript.

## REFERENCES

- Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x
- Blomert, L., and Vaessen, A. (2009). *3DM Differential Diagnostics for Dyslexia: Cognitive Analysis of Reading and Spelling*. Netherlands: Boom Test.
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42:189.
- 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
- 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
- Carey, S., Shusterman, A., Haward, P., and Distefano, R. (2017). Do analog number representations underlie the meanings of young children's verbal numerals? *Cognition* 168, 243–255. doi: 10.1016/j.cognition.2017.06.022
- Cohen, D. J., and Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychon. Bull. Rev.* 18, 331–338. doi: 10.3758/s13423-011-0059-z
- Dackermann, T., Huber, S., Bahnmueller, J., Nuerk, H. C., and Moeller, K. (2015). An integration of competing accounts on children's number line estimation. *Front. Psychol.* 6:884. doi: 10.3389/fpsyg.2015.00884
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.
- 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
- 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., Libertus, M. E., and Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Dev. Perspect.* 7, 74–79. doi: 10.1111/cdep.12019
- Feng, J., Spence, I., and Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychol. Sci.* 18, 850–855. doi: 10.1111/j.1467-9280.2007.01990.x
- 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
- 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
- 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
- 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
- Hutchison, J. E., Lyons, I. M., and Ansari, D. (2018). More similar than different: gender differences in children's basic numerical skills are the exception not the rule. *Child Dev.* doi: 10.1111/cdev.13044 [Epub ahead of print].
- Kimura, D. (1999). *Sex and Cognition*. Cambridge, MA: MIT Press.
- 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.
- 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). 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., 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
- 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
- Lyons, I. M., Ansari, D., and Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain: neural coding of numbers. *Hum. Brain Mapp.* 36, 475–488. doi: 10.1002/hbm.22641
- Lyons, I. M., and Beilock, S. L. (2018). Characterizing the neural coding of symbolic quantities. *NeuroImage* 178, 503–518. doi: 10.1016/j.neuroimage.2018.05.062
- Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256–261. doi: 10.1016/j.cognition.2011.07.009
- Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., and Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. *Dev. Psychol.* 54, 440–457. doi: 10.1037/dev0000445
- 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
- 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
- Merkley, R., and Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: evidence from brain and behavior. *Curr. Opin. Behav. Sci.* 10, 14–20. doi: 10.1016/j.cobeha.2016.04.006
- Nieder, A. (2009). Prefrontal cortex and the evolution of symbolic reference. *Curr. Opin. Neurobiol.* 19, 99–108. doi: 10.1016/j.conb.2009.04.008
- 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
- Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409–424. doi: 10.1016/j.tics.2017.03.005
- Opfer, J., and Siegler, R. (2007). Representational change and children's numerical estimation. *Cogn. Psychol.* 55, 169–195. doi: 10.1016/j.cogpsych.2006.09.002
- Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., and Luwel, K. (2016). Children's use of number line estimation strategies. *Eur. J. Psychol. Educ.* 31, 117–134. doi: 10.1007/s10212-015-0251-z
- Peeters, D., Verschaffel, L., and Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *Br. J. Psychol.* 108, 668–686. doi: 10.1111/bjop.12233
- 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.
- Ramani, G. B., Siegler, R. S., and Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *J. Educ. Psychol.* 104, 661–672. doi: 10.1037/a0028995
- Raven, J. C., Court, J. H., and Raven, J. (1995). "Manual for Raven's progressive matrices and vocabulary scales", in *Coloured Progressive Matrices*, eds J. C. Raven, J. H. Court, and J. Raven (Oxford: Oxford Psychologist Press).
- Roggeman, C., Verguts, T., and Fias, W. (2007). Priming reveals differential coding of symbolic and non-symbolic quantities. *Cognition* 105, 380–394. doi: 10.1016/j.cognition.2006.10.004
- Rouder, J. N., and Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Dev. Sci.* 17, 525–536. doi: 10.1111/desc.12166
- Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement: association between basic numerical abilities and mathematics achievement. *Br. J. Dev. Psychol.* 30, 344–357. doi: 10.1111/j.2044-835X.2011.02048.x
- Sasanguie, D., De Smedt, B., and Reynvoet, B. (2017). 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., 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
- 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.* 89, 1467–1484. doi: 10.1111/cdev.13068
- 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 Braithwaite, D. W. (2017). Numerical development. *Annu. Rev. Psychol.* 68, 187–213. doi: 10.1146/annurev-psych-010416-044101
- Siegler, R. S., Fazio, L. K., Bailey, D. H., and Zhou, X. (2013). Fractions: the new frontier for theories of numerical development. *Trends Cogn. Sci.* 17, 13–19. doi: 10.1016/j.tics.2012.11.004
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Siegler, R. S., and Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *J. Educ. Psychol.* 101, 545–560. doi: 10.1037/a0014239
- Siegler, R. S., Thompson, C. A., and Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cogn. Psychol.* 62, 273–296. doi: 10.1016/j.cogpsych.2011.03.001
- Slusser, E. B., Santiago, R. T., and Barth, H. C. (2013). Developmental change in numerical estimation. *J. Exp. Psychol. Gen.* 142:193–208. doi: 10.1037/a0028560
- Terlecki, M. S., and Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles* 53, 433–441. doi: 10.1007/s11199-005-6765-0
- Thompson, C. A., and Opfer, J. E. (2008). Costs and benefits of representational change: effects of context on age and sex differences in symbolic magnitude estimation. *J. Exp. Child Psychol.* 101, 20–51. doi: 10.1016/j.jecp.2008.02.003
- Vanbinst, K., Ansari, D., Ghesquière, P., and De Smedt, B. (2016). Symbolic numerical magnitude processing is as important to arithmetic as phonological awareness is to reading. *PLoS One* 11:e0151045. doi: 10.1371/journal.pone.0151045
- Van Bon, W. H. J. (1986). *Raven's Coloured Progressive Matrices: Manual of Dutch norms*. Netherlands: Swets Test Services.
- van't Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., and Kroesbergen, E. H. (2016). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychol. Res.* 80, 368–378. doi: 10.1007/s00426-015-0736-z
- Verguts, T., Fias, W., and Stevens, M. (2005). A model of exact small-number representation. *Psychon. Bull. Rev.* 12, 66–80.
- Voyer, D., Voyer, S., and Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* 117, 250–270.

**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 Daker and Lyons. 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.





# Spatial Presentations, but Not Response Formats Influence Spatial-Numerical Associations in Adults

Ursula Fischer<sup>1,2,3\*</sup>, Stefan Huber<sup>3</sup>, Hans-Christoph Nuerk<sup>3,4</sup>, Ulrike Cress<sup>3,4</sup> and Korbinian Moeller<sup>3,4</sup>

<sup>1</sup> Department of Sport Science, University of Konstanz, Konstanz, Germany, <sup>2</sup> Thurgau University of Teacher Education, Kreuzlingen, Switzerland, <sup>3</sup> Leibniz-Institut für Wissensmedien, Tübingen, Germany, <sup>4</sup> Department of Psychology, University of Tuebingen, Tübingen, Germany

## OPEN ACCESS

### Edited by:

Firat Soylu,  
The University of Alabama,  
United States

### Reviewed by:

Maria Grazia Di Bono,  
Università degli Studi di Padova, Italy  
Laura Elizabeth Thomas,  
North Dakota State University,  
United States

### \*Correspondence:

Ursula Fischer  
ursula.fischer@uni-konstanz.de

### Specialty section:

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

**Received:** 14 August 2018

**Accepted:** 04 December 2018

**Published:** 18 December 2018

### Citation:

Fischer U, Huber S, Nuerk H-C,  
Cress U and Moeller K (2018) Spatial  
Presentations, but Not Response  
Formats Influence Spatial-Numerical  
Associations in Adults.  
Front. Psychol. 9:2608.  
doi: 10.3389/fpsyg.2018.02608

According to theories of embodied numerosity, processing of numerical magnitude is anchored in bodily experiences. In particular, spatial representations of number interact with movement in physical space, but it is still unclear whether the extent of the movement is relevant for this interaction. In this study, we compared spatial-numerical associations over response movements of differing spatial expansion. We expected spatial-numerical effects to increase with the extent of physical response movements. In addition, we hypothesized that these effects should be influenced by whether or not a spatial representation of numbers was presented. Adult participants performed two tasks: a *magnitude classification* (comparing numbers to the fixed standard 5), from which we calculated the Spatial Numerical Association of Response Codes (SNARC) effect; and a *magnitude comparison* task (comparing two numbers against each other), from which we calculated a relative numerical congruity effect (NCE), which describes that when two relatively small numbers are compared, responses to the smaller number are faster than responses to the larger number; and vice versa for large numbers. A SNARC effect was observed across all conditions and was not influenced by response movement extent but increased when a number line was presented. In contrast, an NCE was only observed when no number line was presented. This suggests that the SNARC effect and the NCE reflect two different processes. The SNARC effect seems to represent a highly automated classification of numbers as large or small, which is further emphasized by the presentation of a number line. In contrast, the NCE likely results from participants not only classifying numbers as small or large, but also processing their relative size within the relevant section of their mental number line representation. An additional external presentation of a number line might interfere with this process, resulting in overall slower responses. This study follows up on previous spatial-numerical training studies and has implications for future spatial-numerical trainings. Specifically, similar studies with children showed contrasting results, in that response format but not number line presentation influenced spatial-numerical associations. Accordingly, during development, the relative relevance of physical experiences and presentation format for spatial-numerical associations might change.

**Keywords:** spatial-numerical associations, numerical processing, magnitude representation, embodied numerosity, SNARC effect

## INTRODUCTION

Knowledge about numbers and numerical concepts is acquired through interaction with the world around us (e.g., Fischer and Brugger, 2011; Moeller et al., 2012; Myachykov et al., 2014). Although a predisposition to perceive and process magnitudes might be innate or at least present very early in life (e.g., Xu et al., 2005), numerical knowledge is also acquired through physical experiences. Perception of magnitude information, which is often associated with spatial expansion, shapes the way in which magnitudes and numbers are processed (e.g., Fischer and Brugger, 2011; Lindemann and Fischer, 2015). Additionally, physical interaction also seem to play a major role in the acquisition of numerical abilities (e.g., Fischer et al., 2017).

The theoretical account that explains the aforementioned phenomena, also referred to as *embodied numerosity* (Domahs et al., 2010) has received increasing research interest in recent years. Especially finger counting has been described as an example of bodily experiences associated with processing of numerical information and was even argued to lead to a specific finger-based representation of numerical magnitude that persists into adulthood (Fischer and Brugger, 2011; Roesch and Moeller, 2015).

Importantly, there is evidence suggesting that associations between numbers and space can be influenced not only by bodily experiences with fingers but also the whole body (Fischer et al., 2016). In the current study, we not only investigated the effects of bodily movement on the processing of numbers, but also the interplay of movement and visual perception. In the following, we first introduce measures of spatial-numerical associations before giving an overview of the literature on embodied numerosity and embodied trainings. We then summarize previous findings on the interplay of presentation and response in spatial-numerical associations before describing the current study.

## Spatial-Numerical Associations

Numerical magnitude has long been thought to be associated with physical space. This association can either be between numerical and physical extensions (e.g., Henik and Tzelgov, 1982; Siegler and Opfer, 2003; Moeller et al., 2009) or between numbers and a particular direction in space (e.g., Dehaene et al., 1993; see Cipora et al., 2015; Patro et al., 2015a, for this distinction). Regarding spatial directionality, number magnitudes are assumed to be spatially represented along a *mental number line* (see Göbel et al., 2011; Fischer and Shaki, 2014, for reviews). This systematic association of numbers and space seems to develop early in life (e.g., Patro and Haman, 2012; Macchi Cassia et al., 2016; McCrink et al., 2017), and become more and more consolidated until adulthood (Kaufmann et al., 2008; de Hevia and Spelke, 2009).

The mental number line is assumed to be activated automatically whenever number magnitude information is processed (Tzelgov et al., 1992; Dehaene et al., 1993; Rubinsten and Henik, 2005). However, this activation was observed to depend on how relevant number magnitude is for a specific task, and also on how magnitudes are presented and responded to (Nuerk et al., 2005; van Dijck et al., 2009; Fischer et al., 2016). Certain behavioral effects have been established as indicators of

spatial-numerical associations, two of which we considered in the present study.

## The SNARC Effect

One of the most well-known indicators for spatial-numerical associations is the SNARC effect (Spatial Numerical Association of Response Codes, Dehaene et al., 1993). It describes the finding that in Western cultures, small numbers are associated with the left side of space, whereas large numbers are associated with the right side of space (see Wood et al., 2008, for a meta-analysis). Accordingly, when Western participants are asked to respond to smaller numbers with the left hand and to large numbers with the right hand (congruent response direction), they are faster and less error prone than when the response direction is reversed so they have to respond to smaller numbers with the right hand and to larger numbers with the left hand (incongruent response direction, Dehaene et al., 1993). For example, when comparing numbers from 1 to 9 to a fixed standard of 5 in a *magnitude classification task*, responses to the number '2' are made faster with the left than with the right hand, whereas responses to the number '8' are made faster with the right than with the left hand. In the original interpretation, Dehaene et al. (1993) argued that this pattern of results stemmed from an automatic activation of the left-to-right oriented mental number line, with left/right hand responses being either congruent or incongruent with the position of small/large numbers on the mental number line. Alternative accounts, however (e.g., van Dijck and Fias, 2011; Gevers et al., 2010) argue that the SNARC effect does not result from mental number line activation, but rather from working memory processes, or from a verbal coding of the numbers. For example, numbers could be verbally coded as semantically SMALL or LARGE, and the semantic codes could then be associated with the left and right side of space. This verbal coding would be sufficient to illicit a SNARC effect, without the necessity for an explicit processing of the number magnitude (e.g., Gevers et al., 2006c; Proctor and Cho, 2006; Santens and Gevers, 2008; Imbo et al., 2012; see also Schroeder et al., 2017 for a discussion of linguistic influences on the SNARC effect).

## The Numerical Congruity Effect

Another indicator of spatial-numerical associations is the relative numerical congruity effect (NCE) described by Fischer et al. (2016) and based on the congruity effect described by Dehaene (1989). In contrast to the SNARC effect, this effect does not result from changing response assignments for 'smaller' and 'larger' responses. To measure this effect, participants again compare the magnitude of two numbers (e.g., comparing 2–4). They are instructed to respond with the left hand when the target number is smaller than the other number, and with the right hand when it is larger than the other number. However, both of the numbers can vary in size, necessitating an actual *magnitude comparison* between the two numbers rather than a simple classification as smaller or larger than 5. Here, the effect is also calculated by comparing congruent and incongruent responses. However, congruity is not determined by a change in the response direction as for the SNARC effect. Rather, congruity results from a match or mismatch between the absolute size of the number that is

responded to (i.e., small or large) and its relative size to the comparison standard (i.e., smaller or larger). For example, in a congruent comparison, participants have to decide whether the number '2' is smaller or larger than the number '4.' The correct response is 'smaller,' and is made with the left hand, congruently with the position of '2' on the left side of the number line. However, when switching the numbers and comparing '4' to the standard '2,' the *relative* size of the number '4' compared to the number 2 is *larger*, and therefore, a response has to be made with the right hand. However, in the range from 0 to 9, the *absolute* size of '4' is *small*. The resulting incongruence between the absolute and relative magnitude leads to slower and more error-prone responses. The effect can be explained by assuming that for a number to be classified as small within the range 1–9, the mental number line representation of the continuum 1–9 may be co-activated in addition to the magnitude of the to-be-compared numbers. A similar explanation was proposed previously for related effects (i.e., the *semantic congruity effect*, see e.g., Banks et al., 1976; Cantlon and Brannon, 2005). It is therefore possible that the NCE, due to its reliance on activating the entire relevant number range on the mental number line, presents a more direct measure of spatial-numerical associations than the SNARC effect.

## Embodied Numerosity and Embodied Trainings

Recently, spatial-numerical associations have received increasing research interest following numerous studies showing that they are associated with bodily movements (Moeller et al., 2012; Patro et al., 2015b). Indeed, as elaborated on in theories of embodied numerosity (Domahs et al., 2010) bodily movements play an important role in arithmetic and numerical processing, most notably through the use of fingers for counting and representing numbers (e.g., Fischer and Brugger, 2011; Fabbri and Guarini, 2016; Suggate et al., 2017). Most children use their fingers during early numerical development, and the way in which numbers are represented on one's fingers has a substantial impact on the development of spatial-numerical associations (Wasner et al., 2014).

Recent research suggests, however, that bodily movements that interact with spatial-numerical associations generalize from the hands to the whole body (e.g., Fischer, 2003; Schwarz and Müller, 2006; Hartmann et al., 2012; Klein et al., 2014; Shaki and Fischer, 2014). For example, Shaki and Fischer (2014) observed that when participants were asked to make lateral turns to the left or right while walking and generating random numbers, they were more likely to turn left after generating a small number, and to turn right after generating a large number. This finding can be explained by participants associating small numbers with full-body movements to the left and large numbers with full-body movements to the right.

Following the previous studies investigating interactions between numbers and the body, full-body movements have been used to not only measure spatial-numerical associations (Fischer et al., 2016), but also to boost the training success of spatial-numerical trainings in so-called embodied training

approaches (e.g., Dackermann et al., 2017). In most conventional spatial-numerical trainings, participants are trained in a numerical task that also incorporates spatial aspects. For example, children are trained to count numbers which are ordered from left to right or to estimate the position of numbers on a presented number line (e.g., Siegler and Ramani, 2009; Kucian et al., 2011; Sella et al., 2016). The goal of these trainings is to help children understand numerical concepts or to improve their mathematical skills. Trainings that highlight the spatial ordering of numbers are often more beneficial than trainings that do not, as children show more pronounced improvement in the trained tasks but also in untrained transfer tasks (e.g., Siegler and Ramani, 2009). Embodied spatial-numerical training approaches take this concept one step further, as they combine this spatial-numerical task presentation with a spatial full-body response movement. Accordingly, children are trained to respond to a spatial-numerical task with a full-body movement. For example, Fischer et al. (2011) presented kindergartners with a number located on a number line and then asked them to decide whether a second number was smaller or larger. This training was more effective when children responded with their entire body (by jumping to the left for smaller and to the right for larger decisions) than when they responded manually. Children specifically improved more in number line estimation (i.e., they were able to more accurately locate numbers on an empty number line) as well as their understanding of counting principles (i.e., they were better able to count backward or in steps of two). Training concepts such as these were already implemented with different types of training tasks (such as number line estimation) and with different age groups (ranging from kindergarten to second grade). In all previous embodied spatial-numerical trainings studies, a task-relevant full-body movement in accordance with the direction of the mental number line further increased training effects (for overviews see Fischer et al., 2015a; Dackermann et al., 2017).

However, the specific working mechanisms of embodied spatial-numerical trainings are not yet fully understood. Previously, Fischer et al. (2011) argued that in accordance with theories of perception-action integration (e.g., Hommel et al., 2001; Hommel, 2009), the combined spatial features of the full-body response movement and the presentation of a number line increased the activation of the mental number line. This increased activation was then assumed to lead to a deeper processing of the task, in turn increasing training gains (for further training studies see also Link et al., 2013; Fischer et al., 2015b; Dackermann et al., 2016).

While the success of full-body spatial-numerical trainings has been investigated and supported several times, the respective training studies also raised the question of whether it was indeed the combination of full-body movements and spatial presentation of training content that increased training effects, or whether either the spatially distributed presentation along a number line or the full-body response would have sufficed. The underlying working mechanisms of these trainings were first investigated in an experimental study by Fischer et al. (2016), in which different response and presentation formats were compared to measure their influence on the strength of spatial-numerical associations.



As the current study builds upon this previous study, we now describe it in more detail.

## The Interplay of Presentation and Response in Spatial-Numerical Associations

Fischer et al. (2016) first investigated the differential effects of bodily responses on spatial-numerical associations in elementary school children. They expected that a full-body response movement that corresponds to the direction of the mental number line would elicit stronger spatial-numerical associations than a verbal response format. Furthermore, they controlled for the effect of an additional explicit presentation of a number line. In doing so, they evaluated spatial-numerical associations by the two effects described above – the SNARC effect and the NCE. They hypothesized that these effects should be modulated systematically by response and presentation formats. More specifically, they expected the most pronounced effects when full-body responses and the explicit presentation of a number line were combined.

They found that, at least in elementary school children, the strength of spatial-numerical associations was only influenced by response format, but not by the presentation of a number line. In particular, a SNARC effect was observed irrespective of response conditions, whereas the NCE was only observed in conditions requiring physical response movements. Thereby, physical response movements seemingly increased spatial-numerical associations, but only when magnitude processing was necessary as reflected by the NCE in magnitude comparison with a variable standard.

However, while Fischer et al. (2016) differentiated between responses conducted with foot movements and verbal responses, it remained unclear whether maybe a manual response movement as used in typical SNARC experiments would have been sufficient to elicit a NCE. Furthermore, the bodily and verbal responses in the previous study differed in another relevant aspect. While the bodily responses were made horizontally (i.e., to the left and right) to correspond to the horizontal orientation of the presented number line, the verbal responses were made vertically so as not to correspond to the horizontal number line orientation. That is, participants responded by saying ‘up’ and ‘down’ rather than ‘left’ and ‘right.’ This confound between spatial orientation and the modality of the response might have limited the generalizability of the results. Because of these two caveats of the previous study, only limited conclusions could be drawn about whether full-body movement influences spatial-numerical associations. Accordingly, more fine-grained research is necessary to determine whether the degree of bodily movement can influence spatial-numerical associations. Furthermore, spatial-numerical associations keep developing after elementary school age (Ninaus et al., 2017). It is therefore plausible to assume that influences of response and presentation format as investigated by Fischer et al. (2016) may look differently in adults, when spatial-numerical associations are stable and do not need further development. Accordingly, the current study was designed to address these previous issues.

## The Current Study

Measuring both SNARC effect and NCE and building directly on the study by Fischer et al. (2016), we examined the strength of spatial-numerical associations for different types of presentation and response formats in adults. As previously observed by Fischer et al. (2016) in children, we expected that deeper magnitude processing should lead to more pronounced SNARC effects and NCEs.

The extent of bodily movement was varied in three different response formats: Verbal, manual, and full-body responses. Although responses were all spatially oriented (to the left or right), we expected that active bodily movement should increase spatial-numerical associations, whereas they should be smaller in verbal responses as previously observed (Fischer et al., 2016). In accordance with the effects of passive full-body movements on numerical processing (Hartmann et al., 2012), we further expected that spatial-numerical associations should be more pronounced for full-body compared to manual responses, because these full-body responses provide additional vestibular information that is absent in manual responses.

Also in line with previous work, we varied stimulus presentation. In embodied numerical training studies, activation of the mental number line was often additionally enhanced by presenting a number line along with the task (for overviews see Fischer et al., 2015a, 2017). However, a previous experimental study with elementary school children found no differences in spatial-numerical effects whether a number line was presented or not (Fischer et al., 2016). Accordingly, the question whether number line presentation thus leads to more pronounced spatial-numerical associations in addition to the response format has not yet been fully resolved. Therefore, the adult participants in our study also received two different presentation formats. The to-be-compared numbers were either presented along a horizontal number line ranging from 0 to 10 or above each other without a number line.

Finally, we were interested in whether response format and mode of stimulus presentation would interact in affecting spatial-numerical associations. Assuming that both presentation and response format impact spatial-numerical associations, there should be an additive effect on SNARC effect and NCE, with the strongest effects being present when a number line was presented and a full-body movement is required as the response.

Previous results also indicated that the SNARC effect might not only reflect spatial-numerical associations but also aspects of verbal coding. In turn, the SNARC effect should occur regardless of response format. However, the NCE might be more exclusively determined by spatial-numerical associations, which is why we expected it to increase steadily with the extent of the required response movement.

## MATERIALS AND METHODS

### Participants

Prior to testing, we conducted an *a priori* power analysis to determine the necessary number of participants using the

program G\*Power 3.1.9.2 (Faul et al., 2009). We assumed small effect sizes of around  $f = .1$  for both the SNARC effect and NCE, and wanted to acquire a statistical power of 0.90. Accordingly, we entered  $2 \times 3 \times 2 = 12$  measurements for our within-subject design and assumed a strong correlation between our repeated measures of 0.8. The power analysis suggested a sample size of at least 37 participants.

Forty-five university students took part in the study. Out of these, five had to be excluded from the analysis due to missing data. In two cases, the voice key software did not recognize the participants' voice onset correctly, and in three cases, technical difficulties lead to missing data files. Out of the remaining 40 participants (13 male; age:  $M = 21.6$  years,  $SD = 2.9$  years, range = 18–30 years), 35 reported being right-handed. Written informed consent was obtained from participants and the study was approved by the local ethics committee.

Tasks and Effects

To measure SNARC effect and NCE, we used two types of numerical comparison tasks. In both tasks, participants decided whether the magnitude of a target number was smaller or larger than a simultaneously presented comparison standard. To distinguish the target from the standard, the rectangle surrounding the standard was marked by additional cross-shaped lines (see Figure 1).

The paradigms differed with respect to the comparison standard, which was fixed in the SNARC task (*magnitude classification*) and variable in the NCE task (*magnitude comparison*). This difference in comparison standards impacts the relevance of magnitude processing: While in magnitude

classification, a number only has to be classified as small or large, magnitude comparison requires an actual magnitude comparison between the two numbers (Dehaene, 1989).

**Magnitude Classification Task (Fixed Standard)**

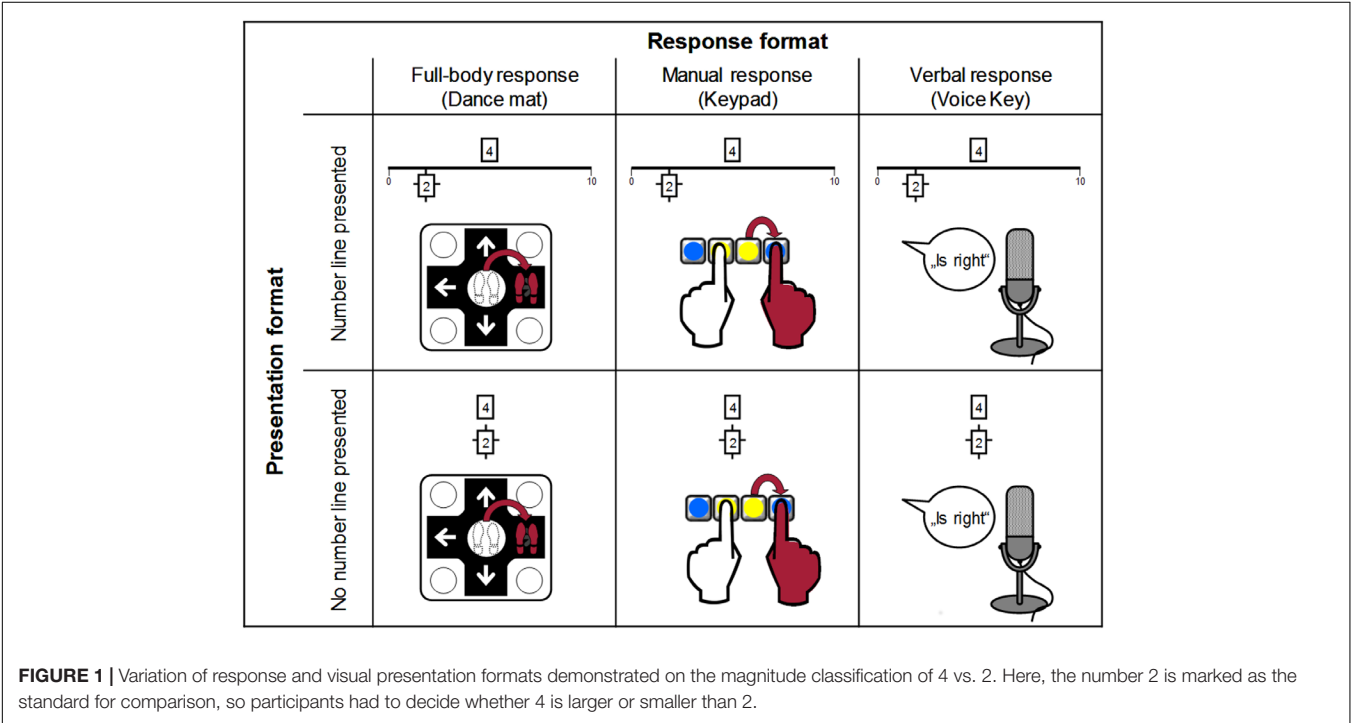
In magnitude classification, numbers had to be compared to the fixed comparison standard 5 (see also Nuerk et al., 2005; Gevers et al., 2006c, 2010). To evaluate the SNARC effect, we varied response direction congruity (congruent vs. incongruent). In the number line congruent direction, participants responded to the left for 'smaller' decisions and to the right for 'larger' decisions, whereas in the number line incongruent direction<sup>1</sup>, they responded to the right for 'smaller' decisions and to the left for 'larger' decisions. The SNARC effect was then calculated by comparing the incongruent and congruent response direction condition (for a similar procedure see Mapelli et al., 2003; Gevers et al., 2006a,b).

In the magnitude classification task, 5 was used as the fixed standard and 1, 4, 6, 9 as targets. All numbers were presented at an equal frequency (each number 12 times per condition) and in random order.

**Magnitude Comparison Task (Variable Standard)**

In magnitude comparison, the comparison standard was varied like the comparison probe from trial to trial (see also Banks et al., 1976; Cantlon and Brannon, 2005). Other than in magnitude classification, participants always responded 'smaller' to the left and 'larger' to the right. To evaluate the NCE, we varied whether

<sup>1</sup>Note that we tested German participants, who grow up with a left-to-right reading/writing direction and therefore in their majority associate larger numbers with the right side and small numbers with left. Congruity might be defined differently in participants from right-to-left reading/writing cultures.



the correct response ('smaller' or 'larger') corresponded to the absolute magnitude (small or large) of the to-be-compared number and thus, its position on the mental number line. For example, compared to the standard 4, the number 2 requires a 'smaller' response to the left. Because within the relevant range of 1–9, 2 is a small number that is located on the left side of the mental number line, this leftward 'smaller' response is congruent with the mental number line position of 2. In contrast, when 4 is compared to the standard 2, this would call for a 'larger' response to the right. Now this response is incongruent with the position of the small number 4 on the left side of the mental number line. Implementing these two types of trials, the NCE was analyzed by comparing incongruent and congruent trials.

The standard was a flexible number in the range between 1 and 9 (excluding 5), and both numbers of a pair were always either smaller than 5 or larger than 5. We used all possible number pairs in the range from 1 to 9, which resulted in a total of 12 number pairs (smaller than 5: 1–2, 1–3, 1–4, 2–3, 2–4, 3–4; larger than 5: 6–7, 6–8, 6–9, 7–8, 7–9, 8–9). Each pair was presented eight times per condition, four of which as a congruent pairing (e.g., 2–4) and four as an incongruent pairing (e.g., 4–2), again in random order.

## Procedure and Apparatus

Participants were tested individually in a university lab. Each participant came in for two sessions that lasted approximately 55 min each. In each session, participants were given the opportunity to take a break in between the different response conditions. To keep experimental conditions as comparable as possible across response conditions, all tasks were presented by projecting them onto a wall in front of participants at a distance of 2.5 m. Tasks were programmed in Java Eclipse and ran on a standard notebook (Fujitsu Siemens Lifebook T 4010).

The three different response formats were implemented using three different types of response media. In the *verbal response condition*, participants responded by speaking their answer into a microphone that was placed on a desktop and adjusted in height for each participant. Participants responded by either saying 'Is left.' (Translated from the German 'Ist links.') or 'Is right.' (German: 'Ist rechts.'). A voice key programmed into the experimental software registered response latencies by detecting the onset of speech, while response accuracy was recorded manually by the experimenter. The verb 'is' was added to allow for the voice key software to capture the actual speech onset analogously for both responses, i.e., without phonemic differences influencing the measured voice onset times.

In the manual response condition, participants were seated at a table and responded on an external numeric keypad with the index fingers of both hands. To avoid the numbers on the keypad interfering with the response and to help participants remember the correct response keys, circular stickers were placed on the keys to cover the numbers. Each trial started with the participant's fingers on two adjacent keys of the keypad, located centrally in the second row from the bottom (keys '2' and '3') and marked with yellow stickers. To respond, participants had to press the key to

the left (key '1') or right ('enter' key) from the starting point of the respective index finger, which were marked with blue stickers (see also **Figure 1**).

In the full-body response condition, we used a digital dance mat (Positive Gaming Impact Dance Pad<sup>2</sup>) with fields arranged in a 3 × 3 layout. Participants responded by hopping from the central field to the right or left field of the dance mat depending on their decision. Both the external keypad and the dance mat were connected to the notebook via USB.

Visual presentation (number line or no number line) was varied by either presenting the comparison standard correctly placed on a number line (endpoints marked 0 and 10) with the to-be-compared number placed centrally above the number line or by presenting both numbers above each other without a number line (see **Figure 1**).

## Design

The experimental manipulations resulted in a 2 × 3 × 2 design for both tasks. For magnitude classification (fixed standard), the factors were response direction (SNARC compatible/incompatible), response format, and presentation format. For magnitude comparison (variable standard), the factors were congruity, response format, and presentation format. Half of the participants started with magnitude classification (fixed standard), while the other half started with magnitude comparison (variable standard). The order of permutations of the factors was balanced between participants. To this end, we generated 2 × 3 × 2 different task sequences and randomly assigned participants to one of them.

In each of the two tasks, participants completed 576 trials. These trials were presented in 12 blocks of 48 trials in magnitude classification (2 response directions, 3 response formats, and 2 presentation formats), and in 6 blocks of 96 trials in magnitude comparison (3 response formats and 2 presentation formats). Note that response direction was varied in blocks in the magnitude classification task. However, congruity in the comparison task was not blocked and varied on a trial by trial basis, as it was determined by the relationship between the presented magnitudes and thus not dependent on a change of response direction.

## Analysis

Prior to analyses, any response times (RT) below or above 3 standard deviations of each participant's individual mean and all RT faster than 200 ms were removed to control for outliers. Only RT for correct responses were analyzed. RT for magnitude classification (fixed standard) and magnitude comparison (variable standard) were then entered into a within-subject repeated measures design. We conducted separate 2 × 3 × 2 (2 *response direction/numerical congruity* × 3 *response formats* × 2 *presentation formats*) repeated-measures analyses of variance for magnitude classification (testing the SNARC effect with a fixed standard) and magnitude comparison (testing the NCE with a variable standard). The presence of a significant SNARC effect/NCE was determined by a main effect of response

<sup>2</sup><http://www.positivegaming.com>

direction/congruity. Any significant effects involving the three-staged factor response format were followed up by pairwise comparisons between the three response formats to determine the origins of the interaction. Analyses were conducted using SPSS 25 (IBM Corp, 2017).

## Data Availability

Datasets are available on request.

## RESULTS

Overall, participants were faster in the magnitude classification task with a fixed standard ( $M = 851$  ms,  $SD = 272$  ms) than in the magnitude comparison task with a variable standard ( $M = 1052$  ms,  $SD = 301$  ms). Because error rates were very low in both tasks (magnitude classification: 4.2%, magnitude comparison: 5.4%), error rates were not analyzed any further.<sup>3</sup>

Figure 2 gives an overview of the mean effects (SNARC effect/NCE) in RT in each condition of both tasks. An overview over raw RT in each condition can be found in the **Supplementary Material**.

### Results Magnitude Classification (Fixed Standard): SNARC Effect

Analyses revealed a significant overall SNARC effect as indicated by a main effect of response direction  $F(1,39) = 36.47$ ,

$p = 0.000$ ,  $\eta_p^2 = 0.48$ . Participants were faster in the SNARC compatible direction (849 ms) than in the SNARC incompatible direction (915 ms). There was also a significant main effect of response format,  $F(2,78) = 388.66$ ,  $p = 0.000$ ,  $\eta_p^2 = 9.09$ ;  $RT_{\text{full body}} = 1165$  ms vs.  $RT_{\text{manual}} = 708$  ms vs.  $RT_{\text{verbal}} = 773$  ms. The full-body movement condition led to slower responses than both the manual and the verbal condition, and responses in the manual condition were faster than in the verbal condition.<sup>4</sup>

Number line presentation also yielded a main effect of RT, as responses were slower when a number line was presented than when it was not,  $F(1,39) = 5.61$ ,  $p = 0.023$ ,  $\eta_p^2 = 0.13$ ;  $RT_{\text{nl presented}} = 889$  ms vs.  $RT_{\text{no nl}} = 875$  ms.

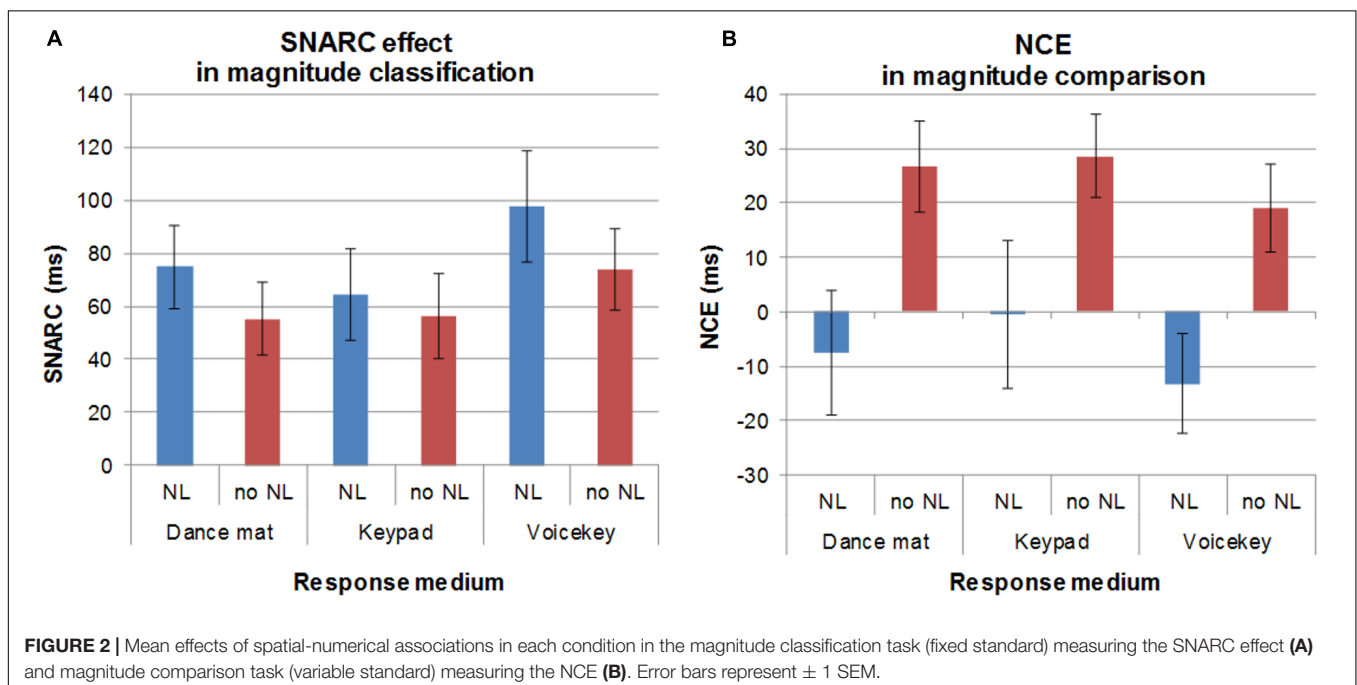
Only the interaction between response direction and presentation format was significant,  $F(1,39) = 4.69$ ,  $p = 0.037$ ,  $\eta_p^2 = 0.11$ . The SNARC effect was more pronounced when a number line was presented than when no number line was presented.

No other interactions reached significance (all  $F < 2.59$ , all  $p > 0.082$ ).

Because we had hypothesized that the SNARC effect should differ between the response formats, but found no significant interaction between response direction and response format to support this hypothesis, we followed up the ANOVA with a Bayesian analysis. This analysis tested the alternative hypothesis that there should be an interaction against the null hypothesis of no interaction between response direction and response format. Using the SPSS\_BAYES\_ANOVA expansion pack for

<sup>3</sup>Comparisons between left- and right-handed participants revealed no differences depending on handedness in SNARC effect [ $t(43) = 1.1$ ,  $p = 0.274$ ] or NCE [ $t(43) = -1.4$ ,  $p = 0.159$ ], and therefore left- ( $N = 5$ ) and right-handers ( $N = 40$ ) were analyzed together.

<sup>4</sup>Because previous studies (see Wood et al., 2008, for an overview) have suggested that the SNARC effect increases with longer response latencies, an additional analysis controlling for this difference in RT between the response formats was conducted and included in the **Supplementary Table 1** for the interested reader.





SPSS Statistics 25.0 (IBM Corp, 2017), we calculated the Bayes factor (alternative/null) for the interaction, which suggested that the data were 0.047:1 in favor of the null hypothesis, or 21.3 times more likely to occur under a model without the interaction than a model including the interaction. According to previously suggested interpretation criteria for the Bayes factor (e.g., Wetzels et al., 2011), this presents strong evidence in favor of the null hypothesis.

## Results Magnitude Comparison (Variable Standard): NCE

In magnitude comparison, we observed no overall significant NCE as indicated by a non-significant main effect of congruity,  $F(1,39) = 2.38$ ,  $p = 0.131$ ,  $\eta_p^2 = 0.06$ ;  $RT_{\text{congruent}} = 1087$  ms vs.  $RT_{\text{incongruent}} = 1095$  ms. Because we had expected a main effect of congruity, we followed this up with a Bayesian analysis. This analysis (alternative/null) revealed that the data were 0.38:1 in favor of the null hypothesis, or 2.62 times more likely under the null than under the alternative hypothesis. This presents only anecdotal evidence in favor of the null hypothesis that there is no overall NCE in the data (e.g., Wetzels et al., 2011), and therefore, the null effect should be interpreted with caution.

As in the magnitude classification task with a fixed standard, there was a main effect of response format,  $F(2,78) = 221.72$ ,  $p = 0.000$ ,  $\eta_p^2 = 0.85$ ;  $RT_{\text{full body}} = 1369$  ms vs.  $RT_{\text{manual}} = 935$  ms vs.  $RT_{\text{verbal}} = 969$  ms. Responses on the dance mat were slower than manual and verbal responses, whereas verbal and manual response condition did not differ in response speed. Presentation format did not yield a main effect,  $F(1,39) = 1.13$ ,  $p = 0.294$ ,  $\eta_p^2 = 0.03$ .

However, congruity interacted significantly with presentation format,  $F(1,39) = 10.58$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.21$ . *Post hoc* comparisons of congruent and incongruent RT indicated that when a number line was presented, there was no significant NCE,  $t(39) = 0.95$ ,  $p = 0.348$ , but there was a significant regular NCE when no number line was presented,  $t(39) = 4.28$ ,  $p = 0.000$ , with incongruent responses (834 ms) being slower than congruent ones (803 ms).

Furthermore, the interaction between response format and presentation format was significant,  $F(2,78) = 3.68$ ,  $p = 0.030$ ,  $\eta_p^2 = 0.09$ . Following this interaction up with *post hoc* pairwise comparisons, we first calculated the differences between the two presentation formats (without vs. with a number line) and compared these across the response formats. There was a significant difference between the full-body and verbal conditions,  $t(39) = 2.55$ ,  $p = 0.015$ , with full-body comparisons being faster with than without a number line (No number line – number line = 18.76 ms), while verbal responses were faster without than with a number line (No number line – number line = –29.07 ms). Furthermore, there was a marginally significant difference between the full-body and manual conditions,  $t(39) = 1.97$ ,  $p = 0.056$ , which again can be explained by the full-body responses showing faster responses with than without a number line (No number line – number line = 18.76 ms) compared to the manual

condition, where responses were faster without than with a number line (No number line – number line = –23.42 ms). No significant difference was observed between the verbal and manual condition,  $t(39) = 0.33$ ,  $p = 0.747$ .

No other interactions reached significance (all  $F < 1.4$ , all  $p > 0.243$ ). Like the SNARC effect, the NCE did not differ significantly between the response formats as indicated by the non-significant interaction between congruity and response format. Again, we therefore followed up the ANOVA with a Bayesian analysis (alternative/null). The analysis revealed that the data were 0.044:1 in favor of the null hypothesis, which corresponds to the data being 22.7 times more likely under a model without the interaction than under a model including the interaction – again indicating strong evidence for the null hypothesis that there is no interaction in the data (e.g., Wetzels et al., 2011).

## DISCUSSION

For the first time, the current study investigated the interplay of response and presentation formats for spatial-numerical associations in adult participants. Following up on previous developmental studies (Fischer et al., 2016), we expected spatial-numerical associations (SNARC effect and NCE) to increase with the extent of left-right physical movements in the response format. Furthermore, we expected that the explicit presentation of a number line should lead to more pronounced spatial-numerical associations. The most pronounced effects were therefore expected for full-body responses in combination with the explicit presentation of a number line. However, our data suggest that these mechanisms may be different in adults compared to children, and that spatial-numerical associations change during development.

Most notably, there were no differences in the strengths of SNARC effect and NCE in the three response conditions. However, unlike in children, adult participants were influenced by the presentation of a number line along with the task, which was not always beneficial. We discuss the theoretical impact of these findings in the following.

## Theoretical Implications

In line with previous work (Fischer et al., 2016), we observed differences in the result patterns for SNARC effect and NCE. In particular, the SNARC effect was again observed in every condition of the magnitude classification task, whereas the NCE was only observed in certain conditions of the magnitude comparison task. However, the SNARC effect differed depending on the presentation format, with number line presentation yielding larger SNARC effects than a presentation without a number line. This influence of number line presentation was not observed in children (Fischer et al., 2016), but seems to indicate an involvement of an underlying mental number line in the occurrence of the SNARC effect in adults.

Regarding the NCE, the picture was more inconsistent, as it was not observed overall, but only when no number line was



presented. However, overall RT did not differ depending on number line presentation. A closer inspection of the marginal means revealed that participants performed at roughly the same speed whenever a trial was incongruent (with NL: 1093 ms; without NL: 1098 ms); and even on congruent trials when a number line was presented (1101 ms). However, when a congruent trial was presented without a number line, response speed increased (1073 ms). Accordingly, the absence of a number line seemed to help participants to solve congruent trials faster. This finding might either indicate that participants did not refer to any spatial-numerical directional representation for solving the congruent trials, and therefore benefited from not having to process redundant visual information. Alternatively, participants might in general rely more on their internal mental number line for the magnitude comparison task, potentially ‘zooming in’ on the relevant section of the number line (i.e., 0–5 when comparing 2 and 4), and can do so more efficiently when they do not have to inhibit an externally presented number line of a non-fitting larger range (i.e., 0–10). However, in the latter case, this should also result in processing advantages for incongruent trials with no number line presentation, which was not supported by the data. Here, future studies would be desirable to further differentiate spatial and numerical aspects of the presentation format.

Another unexpected finding was the interaction between presentation and response format in the magnitude comparison task measuring the NCE. Here, we observed that when participants responded with their entire body, number line presentation led to faster responses compared to a presentation without a number line. However, the opposite was observed for verbal and manual responses, which were descriptively slower with than without number line presentation. While unexpected, this result fits in with previous explanations for why embodied numerical trainings for children have been efficient in the past. For example, Fischer et al. (2015b) as well as Link et al. (2013) observed that combining a presentation of a number line with a full-body response increased the effects of number line estimation trainings compared to trainings that included only number line presentation or a full-body response. A possible explanation for these previous results is that when being presented with a number line and responding with the entire body, this creates an embodied experience of moving along the number line. This fit between the presentation and movement was previously argued to improve training effects and could also account for faster reaction times only in this particular condition in our study.

## Practical Implications for Education and Trainings

Previous studies implementing embodied spatial-numerical trainings suggested that combining spatial-numerical presentation (e.g., a number line) with full-body spatial responses could increase training success (for overviews see Dackermann et al., 2017; Fischer et al., 2017). The first study investigating the underlying working mechanisms of these trainings (Fischer et al., 2016) partially confirmed this interpretation, as in

fourth-graders, response format was more relevant than the presentation of a number line in influencing spatial-numerical associations. However, the current study showed that for adults, the presentation of a number line seemed to play a more prominent role than the response format. Surprisingly, it seemed to hinder rather than to help performance in most conditions.

Within the context of spatial-numerical trainings, the differences in the findings for children and adults might mean that the relevance of each training component (response and presentation) may vary depending on the age of the participants. This possible effect of age should be taken into account when designing future trainings, as older participants might not benefit from an embodied spatial-numerical training in the same way that the young children in previous studies did. To this point, studies on embodied numerical trainings and their underlying mechanisms have only been conducted with children from kindergarten up to fourth grade. It is possible that for children above this age, a full-body response format might not improve training gains, and a presentation of a number line could even hinder training progress. Considering our results, embodied numerical trainings might not even be effective at all for adult participants. However, seeing as the idea behind embodied spatial-numerical trainings is mostly to convey basic numerical competencies, these trainings are not targeted at adult participants. Future studies will be necessary to determine the age at which a full-body response might no longer be adequate. In this vein, longitudinal studies testing the effects of different types of spatial-numerical trainings throughout childhood development would also be informative.

## Limitations and Future Directions

The current study builds on a previous experimental study conducted by Fischer et al. (2016). However, the different age groups investigated mean that the studies are only partially comparable. Because results vary considerably, future studies are needed to close the age gap. In particular, the comparison between manual and full-body responses has not been investigated in children, for whom response format may play a larger role than for adults as indicated by the results of Fischer et al. (2016).

Another aspect to be considered is that task difficulty was possibly not comparable for children and adults. Although the study by Fischer et al. (2016) tested fourth-graders, who should be very familiar with the number range of 0–10, it is reasonable to assume that responses were even more automated for adult participants, and that spatial-numerical effects differed for this reason as well.

A promising avenue for future research could be to test participants across different age groups, while also combining an experimental approach such as the one implemented in the current study with different types of spatial-numerical trainings. Firstly, comparing different age groups within the same paradigm would be informative with regard to what type of training would be most beneficial at what age. Secondly, by measuring spatial-numerical associations before and after

trainings, the relevance of the SNARC effect and NCE as measures of spatial-numerical associations may be further clarified. Furthermore, in case these trainings were to vary in whether they include only a spatial response, a spatial presentation, or both, the effect of each training component on spatial-numerical associations could be distinguished more clearly.

## CONCLUSION

The present findings indicated that adult participants, unlike children, show stable spatial-numerical associations that are independent of the effector with which a task was performed. This suggests that in adults, the strength of spatial-numerical associations is no longer as strongly associated with bodily experiences. Accordingly, while full-body numerical trainings are beneficial for young children, it is possible that trainings for older participants need to take a different approach.

Contrary to previous results of studies with children, visual presentation seemed to play more of a role in adults. However, it was mostly interfering, suggesting that adults' magnitude representations are either (1) more abstract (see e.g., Cipora et al., 2016), such that visuo-spatial perceptual support actually introduces additional interfering information, or (2) more flexible (see e.g., Thompson and Siegler, 2010), such that a fixed number line does not help, but actually hinders flexible zooming in on the number line, as previously shown for other types of spatial-numerical information (Huber et al., 2014).

## REFERENCES

- Banks, W. P., Fujii, M., and Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 435–447. doi: 10.1037/0096-1523.2.3.435
- Cantlon, J. F., and Brannon, E. M. (2005). Semantic congruity affects numerical judgments similarly in monkeys and humans. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16507–16511. doi: 10.1073/pnas.0506463102
- Cipora, K., Hohol, M., Nuerk, H.-C., Willmes, K., Brożek, B., Kucharzyk, B., et al. (2016). Professional mathematicians differ from controls in their spatial-numerical associations. *Psychol. Res.* 80, 710–726. doi: 10.1007/s00426-015-0677-6
- Cipora, K., Patro, K., and Nuerk, H.-C. (2015). Are spatial-numerical associations a cornerstone for arithmetic learning? The lack of genuine correlations suggests no. *Mind Brain Educ.* 9, 190–206. doi: 10.1111/mbe.12093
- 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
- Dackermann, T., Fischer, U., Nuerk, H.-C., Cress, U., and Moeller, K. (2017). Applying embodied cognition: from useful interventions and their theoretical underpinnings to practical applications. *ZDM* 49, 545–557. doi: 10.1007/s11858-017-0850-z
- Dehaene, S. (1989). The psychophysics of numerical comparison: a reexamination of apparently incompatible data. *Percept. Psychophys.* 45, 557–566. doi: 10.3758/BF03208063
- 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

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the 'Ethische Richtlinien der Deutschen Gesellschaft für Psychologie e.V.' (Ethical guidelines of the German Psychological Society) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the 'Lokale Ethikkommission am IWM' (Local ethics committee at KMRC) in Tuebingen, Germany.

## AUTHOR CONTRIBUTIONS

UF, SH, H-CN, UC, and KM conceptualized the study and designed the experiment. SH programmed the experiment. UF conducted the study, analyzed the data, and wrote the first draft of the manuscript. UF, H-CN, UC, and KM wrote the manuscript.

## FUNDING

This work was funded by the German Research Foundation (DFG), grant number CR 110/8-1, granted to UC, H-CN, and KM, supporting UF.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02608/full#supplementary-material>

- Domahs, F., Moeller, K., Huber, S., Willmes, K., and Nuerk, H.-C. (2010). Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition* 116, 251–266. doi: 10.1016/j.cognition.2010.05.007
- Fabbri, M., and Guarini, A. (2016). Finger counting habit and spatial-numerical association in children and adults. *Conscious. Cogn.* 40, 45–53. doi: 10.1016/j.concog.2015.12.012
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Fischer, M. H. (2003). Spatial representations in number processing – evidence from a pointing task. *Vis. Cogn.* 10, 493–508. doi: 10.1080/13506280244000186
- Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260
- 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
- Fischer, U., Klein, E., Dackermann, T., and Moeller, K. (2017). "The physiology of numerical learning: from neural correlates to embodied trainings," in *The Psychology of Digital Learning: Constructing, Exchanging, and Acquiring Knowledge with Digital Media*, eds S. Schwan and U. Cress (Heidelberg: Springer International Publishing), 21–40.
- Fischer, U., Link, T., Cress, U., Nuerk, H.-C., and Moeller, K. (2015a). "Math with the dance mat: on the benefits of embodied numerical training approaches," in *Learning Technologies and the Body: Integration and Implementation in Formal and Informal Learning Environments*, ed. V. Lee (New York, NY: Routledge), 149–163.

- Fischer, U., Moeller, K., Huber, S., Cress, U., Nuerk, C., and Nuerk, H.-C. (2015b). Full-body movement in numerical trainings: a pilot study with an interactive whiteboard. *Int. J. Serious Games* 2, 23–35. doi: 10.17083/ijsg.v2i4.93
- Fischer, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H.-C. (2011). Sensorimotor spatial training of number magnitude representation. *Psychon. Bull. Rev.* 18, 177–183. doi: 10.3758/s13423-010-0031-3
- Fischer, U., Moeller, K., Class, F., Huber, S., Cress, U., and Nuerk, H.-C. (2016). Dancing with the SNARC: measuring spatial-numerical associations on a digital dance mat. *Can. J. Exp. Psychol.* 70, 306–315. doi: 10.1037/cep0000084
- Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., and Fias, W. (2006a). Automatic response activation of implicit spatial information: evidence from the SNARC effect. *Acta Psychol.* 122, 365–377. doi: 10.1016/j.actpsy.2005.11.004
- Gevers, W., Ratinckx, E., De Baene, W., and Fias, W. (2006b). Further evidence that the SNARC effect is processed along a dual-route architecture. *Exp. Psychol.* 53, 58–68. doi: 10.1027/1618-3169.53.1.58
- Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., and Fias, W. (2006c). Numbers and space: a computational model of the SNARC effect. *J. Exp. Psychol.* 32, 32–44. doi: 10.1037/0096-1523.32.1.32
- Gevers, W., Santens, S., Dhooge, E., Chen, Q., Bossche, L., Van den Fias, W., et al. (2010). Verbal-spatial and visuospatial coding of number–space interactions. *J. Exp. Psychol.* 139, 180–190. doi: 10.1037/a0017688
- Göbel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251
- Hartmann, M., Grabherr, L., and Mast, F. W. (2012). Moving along the mental number line: interactions between whole-body motion and numerical cognition. *J. Exp. Psychol.* 38, 1416–1427. doi: 10.1037/a0026706
- Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cogn.* 10, 389–395. doi: 10.3758/BF03202431
- Hevia, M. D., and Spelke, E. S. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition* 110, 198–207. doi: 10.1016/j.cognition.2008.11.003
- Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychol. Res.* 73, 512–526. doi: 10.1007/s00426-009-0234-2
- Hommel, B., Müseler, J., Aschersleben, G., and Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878. doi: 10.1017/S0140525X01000103
- Huber, S., Moeller, K., and Nuerk, H.-C. (2014). Dissociating number line estimations from underlying numerical representations. *Q. J. Exp. Psychol.* 67, 991–1003. doi: 10.1080/17470218.2013.838974
- IBM Corp. (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.
- Imbo, I., Brauwer, J., De 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
- Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L.-B., et al. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex* 44, 376–385. doi: 10.1016/j.cortex.2007.08.003
- Klein, E., Huber, S., Nuerk, H.-C., and Moeller, K. (2014). Operational momentum affects eye fixation behaviour. *Q. J. Exp. Psychol.* 67, 1614–1625. doi: 10.1080/17470218.2014.902976
- 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
- Lindemann, O., and Fischer, M. H. (2015). Embodied number processing. *J. Cogn. Psychol.* 27, 381–387. doi: 10.1080/20445911.2015.1032295
- Link, T., Moeller, K., Huber, S., Fischer, U., and Nuerk, H.-C. (2013). Walk the number line – An embodied training of numerical concepts. *Trends Neurosci. Educ.* 2, 74–84. doi: 10.1016/j.tine.2013.06.005
- Macchi Cassia, V., McCrink, K., Hevia, M. D., de Gariboldi, V., and Bulf, H. (2016). Operational momentum and size ordering in preverbal infants. *Psychol. Res.* 80, 360–367. doi: 10.1007/s00426-016-0750-9
- Mapelli, D., Rusconi, E., and Umiltà, C. (2003). The SNARC effect: an instance of the Simon effect? *Cognition* 88, B1–B10. doi: 10.1016/S0010-0277(03)00042-8
- McCrink, K., Perez, J., and Baruch, E. (2017). Number prompts left-to-right spatial mapping in toddlerhood. *Dev. Psychol.* 53, 1256–1264. doi: 10.1037/dev0000342
- Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., Cress, U., et al. (2012). Learning and development of embodied numerosity. *Cogn. Process.* 13(Suppl. 1), 271–274. doi: 10.1007/s10339-012-0457-9
- Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children's early mental number line: logarithmic or decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006
- Myachykov, A., Scheepers, C., Fischer, M. H., and Kessler, K. (2014). TEST: a tropic. embodied, and situated theory of cognition. *Top. Cogn. Sci.* 6, 442–460. doi: 10.1111/tops.12024
- Ninaus, M., Moeller, K., Kaufmann, L., Fischer, M. H., Nuerk, H.-C., and Wood, G. (2017). Cognitive mechanisms underlying directional and non-directional spatial-numerical associations across the lifespan. *Front. Psychol.* 8:1421. doi: 10.3389/fpsyg.2017.01421
- Nuerk, H.-C., Bauer, F., Krummenacher, J., Heller, D., and Willmes, K. (2005). The power of the mental number line: how the magnitude of unattended numbers affects performance in an Eriksen task. *Psychol. Sci.* 47, 34–50.
- Patro, K., Fischer, U., Nuerk, H.-C. C., and Cress, U. (2015a). How to rapidly construct a spatial-numerical representation in preliterate children (at least temporarily). *Dev. Sci.* 19, 126–144. doi: 10.1111/desc.12296
- Patro, K., Nuerk, H.-C., and Cress, U. (2015b). Does your body count? Embodied influences on the preferred counting direction of preschoolers. *J. Cogn. Psychol.* 27, 413–425. doi: 10.1080/20445911.2015.1008005
- Patro, K., and Haman, M. (2012). The spatial-numerical congruity effect in preschoolers. *J. Exp. Child Psychol.* 111, 534–542. doi: 10.1016/j.jecp.2011.09.006
- Proctor, R., and Cho, Y. (2006). Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychol. Bull.* 132, 416–442. doi: 10.1037/0033-2909.132.3.416
- Roesch, S., and Moeller, K. (2015). Considering digits in a current model of numerical development. *Front. Hum. Neurosci.* 8:10262. doi: 10.3389/fnhum.2014.01062
- 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
- Santens, S., and Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition* 108, 263–270. doi: 10.1016/j.cognition.2008.01.002
- Schroeder, P. A., Nuerk, H.-C., and Plewnia, C. (2017). Switching between multiple codes of SNARC-Like associations: two conceptual replication attempts with anodal tDCS in Sham-controlled cross-over design. *Front. Neurosci.* 11:654. doi: 10.3389/fnins.2017.00654
- Schwarz, W., and Müller, D. (2006). Spatial associations in number-related tasks: a comparison of manual and pedal responses. *Exp. Psychol.* 53, 4–15. doi: 10.1027/1618-3169.53.1.4
- Sella, F., Tressoldi, P., Lucangeli, D., and Zorzi, M. (2016). Training numerical skills with the adaptive videogame “The Number Race”: a randomized controlled trial on preschoolers. *Trends Neurosci. Educ.* 5, 20–29. doi: 10.1016/j.tine.2016.02.002
- Shaki, S., and Fischer, M. H. (2014). Random walks on the mental number line. *Exp. Brain Res.* 232, 43–49. doi: 10.1007/s00221-013-3718-7
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Siegler, R. S., and Ramani, G. B. (2009). Playing linear number board games – but not circular ones – improves low-income preschoolers' numerical understanding. *J. Educ. Psychol.* 101, 545–560. doi: 10.1037/a0014239
- Suggate, S., Stoeger, H., and Fischer, U. (2017). Finger-based numerical skills link fine motor skills to numerical development in preschoolers. *Percept. Mot. Skills* 124, 1085–1106. doi: 10.1177/0031512517727405
- Thompson, C. A., and Siegler, R. S. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychol. Sci.* 21, 1274–1281. doi: 10.1177/0956797610378309
- Tzelgov, J., Meyer, J., and Henik, A. (1992). Automatic and intentional processing of numerical information. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 166–179. doi: 10.1037//0278-7393.18.1.166
- van Dijck, J.-P., and Fias, W. (2011). A working memory account for spatial-numerical associations. *Cognition* 119, 114–119. doi: 10.1016/j.cognition.2010.12.013

- 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
- Wasner, M., Moeller, K., Fischer, M. H., and Nuerk, H.-C. (2014). Aspects of situated cognition in embodied numerosity: the case of finger counting. *Cogn. Process.* 15, 317–328. doi: 10.1007/s10339-014-0599-z
- 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 885 t tests. *Perspect. Psychol. Sci.* 6, 291–298. doi: 10.1177/1745691611406923
- 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.
- Xu, F., Spelke, E. S., and Goddard, S. (2005). Number sense in human infants. *Dev. Sci.* 8, 88–101. doi: 10.1111/j.1467-7687.2005.00395.x
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Fischer, Huber, Nuerk, Cress and Moeller. 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.



# Sex Differences in the Performance of 7–12 Year Olds on a Mental Rotation Task and the Relation With Arithmetic Performance

Marleen van Tetering<sup>1\*</sup>, Marthe van der Donk<sup>1</sup>, Renate Helena Maria de Groot<sup>2</sup> and Jelle Jolles<sup>1</sup>

<sup>1</sup> Faculty of Behavioral and Movement Sciences, Centre for Brain and Learning, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, <sup>2</sup> Welten Institute, Research Centre for Learning, Teaching, and Technology, Open University of the Netherlands, Heerlen, Netherlands, <sup>3</sup> School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands

## OPEN ACCESS

### Edited by:

Sharlene D. Newman,  
Indiana University Bloomington,  
United States

### Reviewed by:

Nicole D. Anderson,  
MacEwan University, Canada  
Laura Piccardi,  
University of L'Aquila, Italy

### \*Correspondence:

Marleen van Tetering  
mvantetering@outlook.com

### Specialty section:

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

**Received:** 24 August 2018

**Accepted:** 14 January 2019

**Published:** 30 January 2019

### Citation:

van Tetering M, van der Donk M,  
de Groot RHM and Jolles J (2019)  
Sex Differences in the Performance  
of 7–12 Year Olds on a Mental  
Rotation Task and the Relation With  
Arithmetic Performance.  
Front. Psychol. 10:107.  
doi: 10.3389/fpsyg.2019.00107

This study evaluates boy-girl differences in 3D mental rotation in schoolchildren aged 7–12 years and the relation to arithmetic performance. A dedicated new task was developed: The Mental Rotation Task – Children (MRT-C). This task was applied to a large sample of 729 children. At the age of 7- to 9-years, a sex difference was found in the number of correct judgments made on the MRT-C. Boys performed better than girls. A closer look at the distribution of boys and girls in this age group showed that boys were overrepresented in the top performance quartile, whereas girls were overrepresented in the lowest performance quartile. A second finding was that higher mental rotation performance was significantly correlated to better mathematical achievement. This finding was done for boys, but not for girls. This correlation underscores the important role that spatial processing plays in mathematical achievement and has implications for school practice.

**Keywords:** 3D mental rotation, childhood, early adolescence, mathematics, STEM

## INTRODUCTION

Mental rotation skills play an important role in achievement in Science, Technology, Engineering, and Mathematics (i.e., STEM, see Wai et al., 2009; Wei et al., 2012; Newcombe and Frick, 2010; Bruce and Hawes, 2015). Moreover, previous studies have reported on the importance of 3D mental rotation skills to school geometry performance at the age of 13 years old (Delgado and Prieto, 2004), to mental mathematics at the age of 15–16 years (Kyttälä and Lehto, 2008) and to algebra at the age of 18–25-years (Tolar et al., 2009). Based on these studies, it can be concluded that there are various ways in which 3D mental rotation skills be wielded throughout mathematics in adolescents and (young) adults. It is now of interest to investigate this link in schoolchildren. This is because there is sufficient evidence that spatial reasoning—including mental rotation—is malleable and susceptible to environmental influences, especially in young children (see also Feng et al., 2007; Zelazo and Carlson, 2012; Uttal et al., 2013). If the 3D mental rotation skills of primary school age children are linked to their mathematical achievement, spatial intervention and enrichment programs could be developed to enhance the development of 3D mental rotation skills and thereby facilitate mathematical achievement.



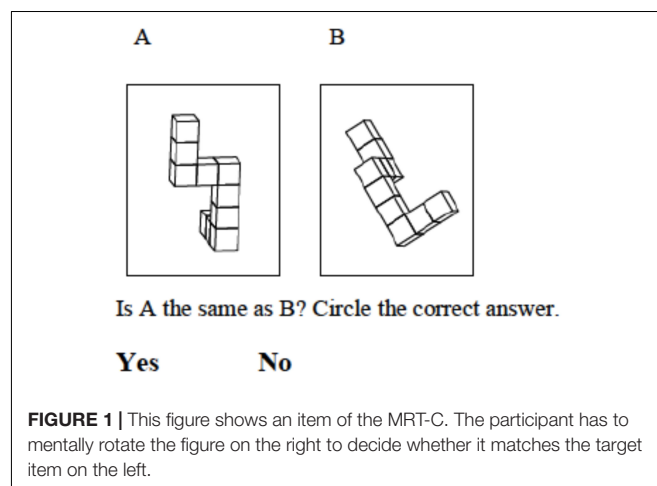
The malleability of spatial skills is also relevant in relation to the well-documented difference between boys and girls in mathematical achievement already at primary school (e.g., Nuttall et al., 2005; Miller and Halpern, 2014; Casey et al., 2015). It may indirectly be the consequence of the preference for spatial play of young boys (Cherney and London, 2006). This preference makes them experienced in spatial skills (Moè, 2016). It gives them a developmental advantage in comparison to girls. The development of spatial abilities – including 3D mental rotation skills – of girls could thus be lagging behind just because they have fewer experiences with spatial play. If these sex differences in spatial skills contribute to differences in the mathematical achievement of boys and girls, intervention programs that stimulate the development of spatial abilities of young girls are promising. These intervention programs may contribute to reducing the well-documented sex differences in successes and achievements in STEM at later ages (Hango, 2013; Miller and Halpern, 2014). This reasoning motivates our study into the relation between 3D mental rotation ability and mathematical achievement in boys and girls. The purpose of this large-scale study was twofold: first, to investigate the contribution of 3D mental rotation to mathematical achievement in 7–12-year old schoolchildren, and second, to investigate sex differences in these participants. More than 700 children participated in the current study.

A dedicated 3D mental rotation task is needed to investigate 3D mental rotation skills in primary school age children. A typical task to evaluate mental rotation requires the participant to compare series of 3D images of objects. The objects may be identical, but rotated around a vertical or horizontal axis or they may be mirror images of each other (Shepard and Metzler, 1971). The participant is asked to determine as quickly as possible which of the images represent the same object but from another rotation (e.g., Peters et al., 1995; Voyer et al., 1995; Hahn et al., 2010; Titze et al., 2010; Hoyek et al., 2011). Large sex differences on such 3D mental rotation tasks are widely reported in adolescent and adult populations (see Voyer et al., 1995). In primary schoolchildren, however, findings of previous studies have given mixed results.

In reviewing the literature that investigated 3D mental rotation skills in schoolchildren, it is important to note that many different tasks have been used. There are substantial differences between these tasks in their procedures and stimuli. Many researchers in young children have used tasks that offer concrete objects (i.e., figures of animals or airplanes) as to-be-rotated stimuli. For instance, Hahn et al. (2010) investigated sex differences in 5-year-old children using colored drawings of animals. Children were asked to indicate whether drawings were identical or mirror-reversed. They found that boys outperformed girls. Another example is the study of Frick et al. (2013) who studied sex differences in the 3D mental rotation skills of 3–5-year-old children. Children saw pairs of asymmetrical ghost figures in seven orientations. One of the ghosts would fit into a hole if rotated right-side up, while the other ghost was its mirror image and would not fit. A disadvantage of the approaches used by Hahn et al. (2010) and Frick et al. (2013) is that they may prompt children to engage primarily in the recognition of object features; children are able to recognize the same figure

as the target figure by comparing the object features of both figures and children are not required to mentally rotate the figures to indicate the correct answer (Hoyek et al., 2011). A second disadvantage of the use of concrete objects as stimuli is that they may elicit an emotional reaction based on the child's positive or negative experiences with that object. It is well-known that emotional reactions can facilitate or hinder memory processing (Christianson and Loftus, 1990; Bradley et al., 1992). If a concrete object elicits a positive emotional reaction, children may store the image more easily into their memory which can facilitate mental rotation of the image. The same accounts for images that elicit negative emotions: this hinders children to store the image into their memory and it is therefore more difficult to mentally rotate the image. For these two reasons, other kinds of stimuli – such as those of the well-established Vandenberg and Kuse Mental Rotation Task (VMRT) – are better suitable to administer 3D mental rotation skills without confounding by object recognition and emotional factors (Vandenberg and Kuse, 1978; Peters et al., 1995).

The VMRT has been used by researchers in schoolchildren. For instance, Titze et al. (2010), Hoyek et al. (2011), and Moè (2018) investigate sex differences in 7- to 12-year-old children. The VMRT requires children to mentally rotate three-dimensional cuboid figures (see Figure 1, Shepard and Metzler, 1971). Various researchers concluded that these figures can reliably be used from early ages onwards (see for instance Örnkloo and von Hofsten, 2007 who presented the figures to 22 months old infants, and Moore and Johnson, 2011 who presented the figures to 3 months old infants). The VMRT asks participants to indicate which two out of four cuboid test figures are rotations of the target figure, rather than mirror versions of it. Both Titze et al. (2010) and Hoyek et al. (2011) reported sex differences in children aged 10 years and older, but not in children under the age of 10 (Titze et al., 2010; Hoyek et al., 2011). These researchers therefore concluded that 10 is the age at which sex differences in 3D mental rotation emerge. Moè (2018), on the other hand, reported sex differences from the age of 8 onwards. Previous research findings are thus inconsistent about the age at which sex differences on the VMRT at first emerge.



Close examination of the VMRT as used by earlier studies (i.e., Titze et al., 2010; Hoyek et al., 2011; Moè, 2018) shows why administering it to young children is problematic. The task may not be comprehensible enough to children under the age of 10 because it is a highly complex task, which requires a high working memory capacity. Working memory is required because (1) the participant needs to remember the task instructions and the target stimulus, (2) to mentally rotate the various alternative stimuli one by one and (3) to remember responses to earlier test stimuli whilst mentally rotating the remaining test stimuli. Next to working memory, this task depends upon several other executive functions (e.g., Anderson et al., 2001; Diamond, 2013; Jolles, 2016). Accordingly, (4) planning and prioritizing are necessary. In addition, (5) high levels of selective attention are needed in order not to be distracted by other options. Finally (6), the participant needs to suppress the tendency to act before thinking and thus have enough impulse control. It can be concluded that the VMRT is a highly complex task for children and involves various executive functions. Tasks that depend heavily on executive functions can be difficult for 8- to 10-year-old children, and even for many 10- to 14-year-old children. The reason is that executive functions are still immature in childhood, as they continue to develop in childhood until at least early adulthood (Diamond, 2013; Jolles, 2016). If the VMRT is too difficult for children aged less than 10, it is possible that the sex difference in performance goes unnoticed. This notion is substantiated by the finding in the study of Hoyek et al. (2011). These authors found low mean performances amongst 7- to 8-year-old children; the mean number of correct responses was similar for the young boys and girls and they performed equally bad on the task. This floor effect could have masked sex differences. We conclude from this body of research that studying mental rotation in young children requires the use of an age-appropriate task that is not too difficult in order to be sensitive to group differences in performance.

Various researchers have therefore modified the VMRT to use it in schoolchildren. Hawes et al. (2015), for example, reduced task difficulty by using tangible figures as to-be-rotated stimuli, instead of the line drawings of 3D cube figures as in the VMRT. Children aged 4–8 years old needed to indicate which out of three figures was identical to the target figure. The results of Hawes et al. (2015) revealed no sex differences in performance. It is notable, however, that this task is still cognitively demanding for young children because it required a comparison between three possible alternatives. Other researchers have therefore used a binary response approach to reduce cognitive demands (e.g., Heil and Jansen-Osmann, 2008; Hahn et al., 2010; Jansen et al., 2013). For example, Casey et al. (2008), reported sex differences in the performances of 6-year-old children on a 3D mental rotation task with a binary response approach. Children had to indicate whether two tangible cuboid 3D figures were the same or not. Boys outperformed girls on this task. Another example of a study that has used a binary response approach is that of Jansen et al. (2013). They investigated sex differences in 3D mental rotation skills in 8 and 10-year-old schoolchildren. In their study, children had to indicate whether two line drawings of cuboid figures were the same or not. In contrast to Casey et al. (2008),

their results revealed no differences between boys and girls in their performances. In fact, they found that the schoolchildren performed beneath chance. Taking the findings of these earlier studies into consideration, it can be concluded that previous studies using modified versions of the VMRT in schoolchildren are inconclusive about the existence of sex differences in 3D mental rotation ability. Our study was therefore carried out to re-investigate the findings of these previous studies. Accordingly, we have modified the VMRT paper-and-pencil test based on the findings of these earlier studies, to make it more suitable for assessing 3D mental rotation in children under the age of 10; the Mental Rotation Task – Children (MRT-C).

In the MRT-C children are asked to indicate whether two stimuli are the same or not. They only have to compare one stimulus with the target, not several, as in the standard VMRT. As our task relies less on executive functions such as working memory, planning and prioritizing and sustained attention, it is easier to apply in young children. In addition to reducing the complexity of the response options, we also reduced the complexity of the stimuli. We limited the stimuli to three-dimensional cuboid figures rotated around a vertical axis by 0 to 180° relative to the target stimulus, whilst the VMRT test stimuli can be rotated around either the horizontal or vertical axis between 0 and 360° (Peters and Battista, 2008). The stimuli for the MRT-C were thus more homogeneous than those for the VMRT and the instructions were easier to understand (Neuburger et al., 2015). This reduced the possibility that children would make procedural mistakes.

In short, there is not enough research to draw conclusions about the age at which the sex gap in 3D mental rotation performance begins to occur (Moè, 2018). Differences between boys and girls in 3D mental rotation skills may contribute to the well-documented sex differences in mathematical achievement that already exists in young children (Miller and Halpern, 2014). They may also contribute to differences between boys and girls in performances and achievement in STEM disciplines at later ages. The aims of the present study were therefore (1) to determine whether there are differences between boys and girls in the performance on the MRT-C in children aged 7–12-years old, and (2) to evaluate the importance of 3D mental rotation ability to mathematical achievement in schoolchildren. We planned a large, cross-sectional study as we wanted to have sufficient power to detect sex differences and to be able to collect information with respect to mathematical achievement at school. Note that previous studies reported an increase in the magnitude of sex differences on 3D mental rotation tasks with age from adolescence onwards (Voyer et al., 1995). It is therefore hypothesized that there are relatively small sex differences in childhood, whereas sex differences in early and later adolescence are more pronounced. A large study sample is needed to detect subtle differences. Our sample therefore consisted of 729 children, and is thereby much larger than that of any previous study (e.g., Voyer et al., 1995; Titze et al., 2010; Hoyek et al., 2011; Jansen et al., 2013). We limited our investigation to children who can be considered to show normal cognitive development; children with evident learning

dysfunction and/or problems in the domain of mental health were excluded.

## MATERIALS AND METHODS

### Participants

The study was part of a large-scale cross-sectional research program called BrainSquare (in Dutch: BreinPlein), which took place in the period of January to June 2016. BrainSquare was aimed at improving knowledge about child-related determinants of learning performance and neurocognitive development of children and young adolescents aged 7 to 12 years (i.e., grades 2 to 6). A total of 1,081 participants were recruited from nine mainstream primary schools in a rural area in the greater Amsterdam region of the Netherlands. Schools were part of the same board and provided roughly equivalent numbers of children from low, middle and high socio-economic status (SES) families. This was done to homogenize our sample with respect to SES. Accordingly, the nine schools were matched on their SES. The SES of the school was established using a composite score that was calculated based on the mean educational levels, incomes, and positions on the labor market of all habitants in the neighborhood of the school in 2016 (Status Scores, 2016). The SES of the schools gives a suitable approximation of the SES of the family in which children grow up in the Netherlands (Central Office for Statistics, 2016). As the study sample included roughly equivalent numbers of children from low, middle and high SES, it is prevented that SES differences between children influenced our main outcomes.

In total,  $N = 1,081$  children participated in the study. Participants were excluded based on the following criteria: (a) skipping or repeating a class ( $n = 231$ ), (b) missing data about the participants age ( $n = 46$ ) or sex ( $n = 4$ ), (c) missing data on the mental rotation task ( $n = 54$ ), and (d) unreliable data because the child did not understand the task-instructions ( $n = 17$ ). By excluding the participants that skipped or repeated a grade, we homogenized the sample by including only the typically developing participants in each grade. All children in the sample can be considered healthy, and the sample is a representative selection of normal and healthy children in primary school. The final sample consisted out of  $n = 729$  individuals (48.8% girls). Of these participants, 137 subjects were in grade 2 (50.4% girls;  $Mage = 7.75$ ,  $SE = 0.02$ ), 123 participants were in grade 3 (41.5% girls;  $Mage = 8.82$ ,  $SE = 0.03$ ), 156 participants were in grade 4 (52.6% girls;  $Mage = 9.84$ ,  $SE = 0.02$ ), 132 participants were in grade 5 (47.7% girls;  $Mage = 10.76$ ,  $SE = 0.03$ ), and 181 participants were in grade 6 (50.3% girls;  $Mage = 11.88$ ,  $SE = 0.03$ ). An analysis of variance (ANOVA) revealed that the average age of boys and girls in each grade did not significantly differ between the sexes ( $p$ -values between 0.28 and 0.96).

For statistical analyses in which sex differences were investigated, the participants were analyzed in two age groups: one group consisting of 416 participants with a mean age of 8.9 years (grades 2–4; 46.2% girls; age range = 7.3–10.4,  $SE = 0.05$ ) and one group consisting of 313 participants with a mean age of 11.4 years (grades 5 and 6; 50.3% girls; age range = 10.3–12.9,  $SE = 0.04$ ). Again, ANOVA revealed that the average age of girls

and boys did not significantly differ in the younger age group [ $F(1,414) = 0.06$ ,  $p = 0.81$ ,  $\eta_p^2 = 0.00$ ], and in the older age group [ $F(1,311) = 0.16$ ,  $p = 0.69$ ,  $\eta_p^2 = 0.00$ ].

To evaluate the importance of mental rotation to mathematical achievement, all children were included with complete data on a standardized mathematical achievement test. This included 121 (49.6% girls) participants in grade 2, 108 (39.8% girls) participants in grade 3, 129 (50.4% girls) participants in grade 4, 110 (45.5% girls) participants in grade 5 and 121 (51.2% girls) participants in grade 6.

### Procedure

First, the collaborating schools agreed to include the testing procedure into their regular school schedule. Then, parents or caregivers (referred to as caregivers in the rest of the paper) of the participating schools received an information letter about the study and gave written informed consent. Children gave verbal consent to participate. Participation was voluntary. All caregivers were informed that no personalized data would be used in the analyses and that no personalized results would be obtained, since all data were assembled on group level. The Ethical Committee of the Faculty of Behavioural and Movement Sciences of the Vrije Universiteit Amsterdam approved the study protocol.

The children were tested at their own school during normal class time. Questionnaires and neuropsychological tests were administered by means of group administration. This was procedurally identical for every class. A maximum of 30 children was tested together in the classroom. Administration of the total protocol took approximately 60 min. All schools were tested within 3 weeks. Tests were administered by the same two neuropsychologists. One of them gave instructions to the participants and kept track of time. The other walked around in the classroom to assist the school teacher with procedural problems. Additionally, the teacher supported with task administration and kept order in the class.

The data analyzed in the study are part of a larger study protocol consisting of eight neuropsychological tests. Participants first filled in their sex, handedness and their date of birth. The mental rotation task was the sixth task within this protocol and took about 5 min to administer. After task administration, data on the mathematical achievement of each individual child were provided by the school.

## Measures

### The Mental Rotation Task – Children

Participants had to solve the Mental Rotation Task – Children (MRT-C), which is a newly made, modified version of the VMRT. The VMRT is a well-established and frequently used task to administer mental rotation ability (Vandenberg and Kuse, 1978; Peters et al., 1995). Both the VMRT and the MRT-C have a similar experimental approach. They are both paper-and-pencil tests that use the 10-block, three-dimensional cuboid figures (i.e., originally introduced by Shepard and Metzler, 1971).

The MRT-C consists of 26 items of three-dimensional-objects, with one reference figure on the left and one figure on the right (see **Figure 1**). All items are derived from the original VMRT and have therefore proven to be valid to assess mental



rotation ability (Vandenberg and Kuse, 1978). The total test was divided into two sets, each containing 13 items. Only figures with rotations in space ranging from 0 to 180° around the vertical axis were selected. The participants had to mentally rotate the target figure and indicate whether the figure on the right matched the reference figure. Earlier studies have proven that this approach (two-answer approach) can validly be used in our age-group (Heil and Jansen-Osmann, 2008; Hahn et al., 2010; Jansen et al., 2013). Participants thus needed to answer a question with binary answer approach: yes or no. All items had a similar difficulty level. They were semi-randomly distributed over the two trials based on their rotation. Furthermore, it was prevented that items did or did not match the target item more than three times in a row to control for answer tendencies (individuals answer yes because this was the answer three times in a row). Finally, split-half reliability was checked for each set and revealed that the first half of the set was as difficult as the second half of the set.

The task-instructions of the MRT-C were explained classically by the researchers using an example item. Then, participants were instructed to solve another item themselves. The answers given by the participants were checked for their accuracy by the researchers. It was then asked whether the task-instructions were completely understood. Each set consisted of five pages. Three items were presented on one page in a booklet (sized 210 by 297 mm). The last page of each set contained only one item. Participants were allowed 2 min to complete each set; a short pause of approximately 1 min was given in between. This pause was devised to reduce possible mental fatigue effects. All participants received the same items in the same order. Credit was given for each item that was correctly marked within the 2 min. Total score for an individual participant could thus range from 0 to 26. Also, the number of mistakes was counted for each individual.

The test had a good split-half reliability (Pearson correlation = 0.60,  $p < 0.01$ ). Only 0.7% of the children received a score of 2 on the test (this was the lowest score obtained), and 0.5% of the children received a score of 26 out of a possible 26 (no participants in grade 2, 1 participant in grade 3, 1 participant in grade 4, no participants in grade 5 and 3 participants in grade 6). These findings indicate that there were no floor or ceiling effects.

### Mathematical Achievement: The Cito Test

Mathematical achievement was assessed with a nationally used paper-and-pencil achievement test, which is standardized and norm-referenced in the Netherlands. This test has been developed by the Dutch Standard Central Institute for Test Development [i.e., in Dutch: Centraal Instituut voor Toetsontwikkeling (Janssen et al., 2010)]. The Dutch Cito mathematics test was used to assess mathematical abilities (Janssen et al., 2010). Participants fill out their answers on a piece of paper. The test took 40–45 min to administer. In grades 3 to 6, the following math skills are covered in the test: (a) number and number relations; (b) addition and subtraction; (c) multiplication and division; (d) measuring (e.g., weights, length, surface, time). From grade 4, (e) percentages and fractions are also covered.

The internal consistency of the Cito mathematics test as a measure of reliability is reported to be high (i.e., for grades 3–6 it ranges from 0.91 to 0.97, see Janssen et al., 2010). The validity of the Cito mathematics test is considered to be high as well since (1) calibration research showed that the differences in participant performance could be explained by one unidimensional concept, (2) similar abilities that were measured with other subparts of the Cito mathematics test were highly correlated, and (3) participants' performances on the Cito mathematics test was predictive for performance on the following Cito test.

In the present study, the “skill-scores” (i.e., translated from the Dutch “vaardigheidscores”) was used as a measure for cognitive performance. These scores are known to improve over the years and are useful in monitoring the progression on each Cito test (Janssen et al., 2010). There are two different test moments for each grade, one regularly administered halfway through the year (January) and one around June. In this study, we used the Cito test results obtained in January 2016.

### Statistical Analyses

All analyses were performed using SPSS version 23. Eta squares were reported as a measure for effect sizes. A total of seven analyses were performed. At first, Pearson correlations were calculated between the Cito mathematical achievement and MRT-C performance in each grade. Secondly, Pearson correlations were calculated for boys and girls separately in each grade. Thirdly, it was investigated whether boys and girls differed in their mathematical achievement per grade using separate one-way analyses of variance (ANOVAs). Modified Hochberg correction was used to control for multiple testing issues; a  $p$ -value of  $<0.01$  was considered statistically significant.

Fourthly, two (age group: younger aged 7- to 9-years old vs. older participants aged 10- to 12-years old)  $\times$  two (sex: boys vs. girls) ANOVAs were performed with MRT-C performance (total number of correctly identified items) as dependent variable. A  $p$ -value of  $<0.05$  was considered statistically significant. Because of the significant interaction between grade and sex, *post hoc* one-way ANOVAs were performed to investigate sex differences in each age group separately. Modified Hochberg correction was used to control for multiple testing issues when assessing sex differences in the two separate age groups. According to this correction, a  $p$ -value of  $\leq 0.04$  was considered critical for assigning statistical significance (Rom, 2013). Then, to investigate more precisely at what age possible sex differences emerge, *post hoc* one-way ANOVAs were performed to investigate sex differences in each study grade. Again, Modified Hochberg correction was used to control for multiple testing issues; a  $p$ -value of  $<0.01$  was considered critical for assigning statistical significance (Rom, 2013).

The fifth analyses were performed to take a closer look at the distribution of boys and girls in the overall sample. MRT-C performance was divided into quartiles ranging from lowest to highest performances (according to a procedure published in Dekker et al., 2013). This was done per grade to control for the age effect which was needed, because MRT-C performance was expected to improve with grade. These analyses provide more insight into the distribution of boys and girls in a group

of low, medium, good, and excellent performers. This reflects a real-life situation, since each class includes performers of various levels.

The sixth analyses were performed using one-way ANOVAs to investigate whether boys and girls (independent variable) differed in their total number of mistakes on the MRT-C (dependent variable) per age group. These analyses were performed to control for the possibility that boys performed better because they prioritized speed above accuracy and thereby achieved a higher number of correct responses because they have been guessing the solutions to some items. According to the Modified Hochberg correction that was used to control for multiple testing issues; a  $p$ -value of  $\leq 0.04$  was considered critical for assigning statistical significance (Rom, 2013).

## RESULTS

### Correlations Between Mathematical Achievement and MRT-C Performance

In the total study population, MRT-C performance was significantly correlated to performance on the mathematical achievement test in grades 2–5 (see Table 1).

The correlation was then investigated for boys and girls separately. Results revealed that MRT-C performance was significantly correlated to mathematical achievement of boys in grades 2–5. For girls, MRT-C performance was significantly correlated to mathematics achievement in grade 2.

### Sex Differences in Mathematical Achievements

Differences in the mathematical achievement between boys and girls were investigated per grade. Results of one-way ANOVAs revealed significant differences in the mean mathematical achievement between boys and girls in grade 2 [ $F(1,119) = 8.76$ ,  $p < 0.01$ ,  $\eta^2 = 0.07$ ] and grade 4 [ $F(1,127) = 9.35$ ,  $p = 0.03$ ,  $\eta^2 = 0.07$ ]. Mean performances of boys (grade 2:  $M = 175.2$ ,  $SD = 29.6$ ; grade 4:  $M = 90.5$ ,  $SD = 10.9$ ) were higher than that of girls (grade 2:  $M = 159.8$ ,  $SD = 27.5$ ; grade 4:  $M = 83.9$ ,  $SD = 13.6$ ). Mean difference in mathematical achievement between boys and girls approaches significance in grade 5 [ $F(1,108) = 4.46$ ,  $p = 0.04$ ,  $\eta^2 = 0.04$ ] (see Table 2). Mean mathematical achievement of boys ( $M = 105.4$ ,  $SD = 11.2$ ) was higher than that of girls ( $M = 100.3$ ,  $SD = 13.9$ ).

**TABLE 1 |** Pearson correlations between MRT-C performance and mathematical achievement.

	<i>N</i> (boys/girls)	<i>r</i> Total	<i>r</i> Boys	<i>r</i> Girls
Grade 2	121 (61/60)	0.38**	0.27*	0.39**
Grade 3	108 (65/43)	0.23*	0.26*	0.16
Grade 4	129 (64/65)	0.19*	0.13	0.15
Grade 5	110 (60/50)	0.26*	0.31*	0.17
Grade 6	121 (59/62)	0.16	0.117	0.22

\* $p < 0.05$ , \*\* $p < 0.01$ .

**TABLE 2 |** Differences between boys and girls in mathematical achievement per grade.

	Boys	Girls	$\eta^2$	$p$ -Value
	Mean (SD)	Mean (SD)		
Grade 2	175.2 (29.6)	159.8 (27.5)	0.07	< 0.01*
Grade 3	154.2 (76.3)	141.4 (71.4)	0.01	0.38
Grade 4	90.5 (10.9)	83.9 (13.6)	0.07	< 0.01*
Grade 5	105.4 (11.2)	100.3 (13.9)	0.04	< 0.04*
Grade 6	114.7 (9.9)	115.5 (11.3)	< 0.01	0.69

\* $p < 0.05$ .

### Sex Differences in Younger and Older Participants in Mental Rotation

Differences between boys and girls, younger and older children, and the possible interaction between sex and age group on MRT-C performance were investigated. Table 3 presents the number of correct substitutions on the MRT-C by age group and sex. Results revealed significant main effects of sex [ $F(1,725) = 14.80$ ,  $p < 0.01$ ,  $\eta_p = 0.02$ ] and age group [ $F(1,725) = 92.28$ ,  $p < 0.01$ ,  $\eta_p = 0.11$ ] on MRT-C performance. Boys ( $M = 16.6$ ,  $SE = 0.26$ ) showed better performance than girls ( $M = 15.2$ ,  $SE = 0.26$ ), and the older participants ( $M = 17.8$ ,  $SE = 0.26$ ) showed better performance than the younger participants ( $M = 14.5$ ,  $SE = 0.23$ ). The interaction between sex and age-group on MRT-C performance was significant as well [ $F(1,725) = 5.22$ ,  $p = 0.02$ ,  $\eta_p = 0.01$ ], indicating that the difference in the performance of boys and girls is different in the older age group than in the younger age group.

Because of the significant interaction, *post hoc* analyses were performed to investigate sex differences within each age group. Results showed an effect of sex on MRT-C performance in the younger age group [ $F(1,414) = 22.01$ ,  $p < 0.01$ ,  $\eta_p = 0.05$ ], but not in the older age group [ $F(1,311) = 1.06$ ,  $p = 0.30$ ,  $\eta_p = 0.00$ ]. More specific, younger boys ( $M = 15.5$ ,  $SE = 0.33$ ) outperformed younger girls ( $M = 13.4$ ,  $SE = 0.31$ ), whereas in the older age group boys and girls performed equally.

### Post hoc Analyses: Sex Differences per Grade

*Post hoc* analyses were conducted in which sex differences on MRT-C performances were investigated in each grade separately. Results showed an effect of sex on MRT-C performance in grade 2 [ $F(1,135) = 9.15$ ,  $p < 0.01$ ,  $\eta_p = 0.06$ ] and in grade

**TABLE 3 |** Mean performance on the MRT-C for younger and older participants, and for boys and girls.

	Total sample	Boys	Girls		
	<i>M</i> (SE)	<i>M</i> (SE)	<i>M</i> (SE)	$\eta_p$	$p$ -Value
Total sample	15.9 (0.19)	16.6 (0.26)	15.2 (0.26)	0.02	0.01*
Younger	14.5 (0.23)	15.5 (0.33)	13.4 (0.31)	0.05	<0.01*
Older	17.8 (0.26)	18.1 (0.40)	17.5 (0.34)	0.00	0.30

\* $p \leq 0.01$ .



4 [ $F(1,154) = 11.82, p < 0.01, \eta_p = 0.07$ ], to the advantage of boys. The sex-difference on MRT-C performance approached significance in grade 3 [ $F(1,122) = 4.61, p = 0.03, \eta_p = 0.03$ ], to the advantage of boys. No sex differences were found in grade 5 [ $F(1,130) = 0.77, p = 0.38, \eta_p = 0.01$ ], and 6 [ $F(1,179) = 0.44, p = 0.51, \eta_p = 0.00$ ]. Means, standard errors,  $p$ -values and effect sizes are presented in Table 4.

## Distribution of Boys and Girls in the Total Study Population

Analyses were performed to take a closer look at the distribution of boys and girls in the overall sample. The distribution of boys and girls differed significantly between the quartiles [ $\chi^2(3) = 21.87, p < 0.01$ ]. It appeared that students in the highest quartile were predominantly boys (boy: girl ratio = 2: 1; boys:  $z = 2.7$ ). There were no significant differences in the boy: girl ratio in the first (boy: girl ratio = 3:4; boys:  $z = -1.3$ ), second (boy: girl ratio = 1: 1; boys:  $z = -0.6$ ) and third quartiles (boy: girl ratio = 3: 4; boys:  $z = -1.2$ ) (see Figure 2).

## Distribution Boys and Girls in the Younger and Older Age Groups

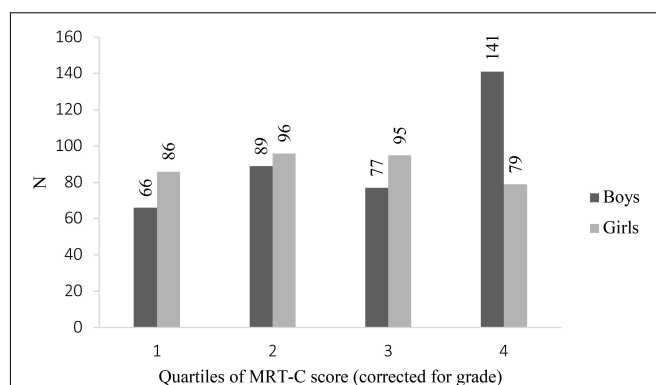
The distribution of boys and girls was investigated in the younger and older age group. Within the younger age group, we found that the relative number of boys and girls significantly differed between the quartiles [ $\chi^2(3) = 19.052; p < 0.01$ ]. It appears that students within the lowest quartile were predominantly girls (boy: girl ratio = 2:3; boys:  $z = -2.8$ ), and within the highest quartile were predominantly boys (boy: girl ratio = 7: 3; boys:  $z = 2.4$ ). There were no significant differences in the boy: girl ratio for the second (boy: girl ratio = 1: 1; boys:  $z = -1.0$ ) and third (boy: girl ratio = 1: 1; boys:  $z = -0.8$ ) quartiles.

Within the older age group, we found that the differences in the distribution of boys and girls approached significance [ $\chi^2(3) = 7.159; p = 0.067$ ]. It appeared that students within the third quartile were predominantly girls (boy: girl ratio = 2: 3; boys:  $z = -2.1$ ), and within the highest quartile were predominantly boys (boy: girl ratio = 3: 2; boys:  $z = 2.3$ ). There were no significant differences in the boy: girl ratio for the first (boy: girl ratio = 1: 1; boys:  $z = -0.2$ ) and second (boy: girl ratio = 1: 1; boys:  $z = -0.3$ ) quartiles.

**TABLE 4 |** Mean performance on the MRT-C and results of the analyses for boys and girls per grade.

	Boys	Girls	$\eta_p$	$p$ -Values
	M (SE)	M (SE)		
Grade 2	13.5 (0.50)	11.4 (0.52)	0.06	<0.01*
Grade 3	15.8 (0.60)	13.8 (0.66)	0.03	<0.03
Grade 4	17.1 (0.53)	14.8 (0.41)	0.07	<0.01*
Grade 5	17.6 (0.61)	16.8 (0.58)	0.01	0.38
Grade 6	18.4 (0.53)	18.0 (0.41)	0.00	0.51

\* $p$ -Value  $\leq 0.05$ .



**FIGURE 2 |** Sex differences in MRT-C performance in the total population divided over quartiles. Quartile 1 = 25% lowest MRT-C scores; quartile 4 = 25% highest MRT-C scores. \* $p$ -Value  $\leq 0.01$ .

## Sex Differences in the Number of Mistakes

Additional analyses were performed to investigate whether boys and girls differed in their total number of mistakes. Results showed that within the younger group, girls ( $M = 7.1, SE = 0.29$ ) made significantly more mistakes than boys ( $M = 6.0, SE = 0.34$ ),  $F(1,414) = 6.10, p = 0.01, \eta_p = 0.02$ . In the older group, no significant difference in the total number of mistakes was found between boys ( $M = 4.8, SE = 0.31$ ) and girls ( $M = 5.2, SE = 0.29$ ),  $F(1,311) = 0.69, p = 0.41, \eta_p = 0.02$ .

## DISCUSSION

This aims of this study were (1) to investigate the correlation between 3D mental rotation and mathematical achievements in 7–12-year-old children, and (2) to investigate whether sex differences in 3D mental rotation were present before the age of 10 years. The MRT-C has been developed for investigating 3D mental rotation performance in children below the age of 10 years. Results revealed that MRT-C performance was positively correlated to higher mathematical achievement, especially for boys. Moreover, there were differences between boys and girls in their mathematical achievements. Boys performed better than girls. The same was found with respect to sex differences on the MRT-C; boys performed better than girls. This sex difference was confined to the younger age group (aged 7–10 years old). Major strength of our study was its large sample size which enabled us to detect this relatively small, but substantial difference in contrast to earlier studies that were much smaller and therefore unable to detect this difference (for instance, Hoyek et al., 2011 investigated sex differences in 22 and 66 participants, and Titze et al., 2010 investigated sex differences in 95 participants). The importance of this finding was substantiated by inspection of the sex distribution of performance in the younger group. Boys were overrepresented in the top performance quartile, whilst the lowest quartile was predominantly made up of girls. The same was found in the older group: there were more boys than girls in the top performance

quartile. Finally, analyses showed that girls made more mistakes than their male peers at the ages of 7- to 10-years, but not at the ages of 10- to 12-years. These analyses were performed to control for the possibility that boys performed better because they prioritized speed above accuracy and achieved a higher number of correct responses because they guessed the answers to some items. This alternative hypothesis was not supported by the data as boys made fewer mistakes than girls. These findings substantiate the notion that boys are better in the task than girls at the age of 7–10 years.

The important finding that 3D mental rotation performance was positively correlated to mathematical achievements in 7–11-year-olds is in line with that of earlier studies in older participants. These studies showed that 3D mental rotation skills are involved in various aspects of mathematics. For instance, 3D mental rotation skills are involved in school geometry when visualizing the lengths of lines or the size of in-depth-figures (Delgado and Prieto, 2004). They are also involved in mental mathematics while holding multiple simultaneous representations of numbers into mind (Kytälä and Lehto, 2008; Thompson et al., 2013). Our finding that young boys are better in 3D mental rotation than girls the same age indicates that boys could be better in visualizing and thinking about representations of numbers into their minds than girls at early ages. This could be beneficial for their mathematical achievements because our results showed that mental rotation performance was significantly correlated to mathematical performance especially in boys. Our finding is substantiated by that of Frick (2018). She reported better mental transformation skills, particularly the ones requiring a high level of spatial flexibility and a stronger sense for spatial magnitudes in boys than in girls. She also found that these skills were beneficial for mathematical performance. This is an important finding when it comes to improving the mathematical achievements of girls. It could explain why boys generally outperform girls in mathematical related fields, as has extensively been reported (e.g., Miller and Halpern, 2014). An interesting implication of our findings is that children should be stimulated to get experience in spatial information processing (mental rotation and other spatial skills) as this can aid in the development of skills that are important for mathematical thinking. It can be envisaged that this applies not only to boys but also to girls; our finding suggest that mathematical achievements of girls could be improved by practicing their spatial abilities. This suggestion should be evaluated in controlled intervention experiment. Moreover, for future research it would be relevant to investigate whether alternative strategy use could be a source of sex differences on MRT-C performance in this age-group. In an adult population, for instance, Boone and Hegarty (2017) found that males outperformed females when items were structurally different so that mental rotation was not necessary. They also found that when all foils were structure foils and participants were instructed to look for structure foils, the significant sex difference was no longer evident. Their findings therefore indicate that there are sex differences in strategy use (males look for structure foils and females do not) that contribute to the sex difference in mental rotation performance. It would now be interesting to investigate

whether this sex difference already exists in young children. This would indicate that mental rotation performance of girls could improve if they learn more efficient strategies.

With respect to the task used, our findings support the notion that the original VMRT is too complex for young children. The task could therefore be insensitive to sex differences. This hypothesis is supported by the fact that the adjusted VMRT (i.e., the MRT-C) is less complex in both the answer approach and the nature of the stimuli used. In contrast to studies using the VMRT (Titze et al., 2010; Hoyek et al., 2011; Hawes et al., 2015), our study did reveal sex differences in 7–10-year-old children. This finding is in line with that of Casey et al. (2008), who used the same answer approach in their task as that of the MRT-C (i.e., their approach was also binary). Because of this simplified answer approach, MRT-C performance is less dependent on executive functions such as working memory, planning and prioritizing and selective attention than performance on the VMRT. It appears that this is important for research in young children given the existence of individual differences in executive functions in 7–12-year-old children (e.g., Anderson et al., 2001; Diamond, 2013; van Tetering and Jolles, 2017). For instance, there is evidence that the child's sex is a relevant factor contributing to individual differences in executive functions (see van Tetering and Jolles, 2017). When difficult tasks – such as the VMRT – are used to assess 3D mental rotation, sex differences in executive functions may interfere with task performances. This unwanted contamination of performance on the target skill can be avoided by using a more straightforward task such as the newly developed MRT-C.

An important strength of the MRT-C is the use of three-dimensional cuboid figures. These stimuli are unlikely to elicit emotional reactions that could influence mental rotation performance. This is one of the reasons that these figures have been used in many earlier studies on sex differences in 3D mental rotation ability in primary school age children (e.g., see Voyer et al., 1995; Titze et al., 2010; Hoyek et al., 2011). Another reason why previous studies used these figures is that there is much evidence that participants actually mentally rotate stimuli of this type into an upright position in order to determine whether pairs of stimuli are identical or mirror images (Hoyek et al., 2011). Moreover, Hawes et al. (2015) concluded that cuboid figures can be used by 4-year-old children. They showed that these children performed above chance in their task using these figures. Cuboid figures are thus highly appropriate stimuli to administer mental rotation ability, and they are useful in young children.

An additionally relevant finding of this study pertains to the fact that sex differences in 3D mental rotation are present in the best performing older children: boys are overrepresented in the upper performance quartile, whereas there was no sex difference in MRT-C on a group level in early adolescents aged 10- to 12-years old. This finding implies that sex differences are especially present in the extreme performance groups including children with excellent 3D mental rotation skills. For instance, if sex differences are investigated in the total study population, substantial sex differences in the extreme performance groups are canceled out by the smaller sex differences in the average performance groups. This may explain why Jansen et al. (2013)

did not find differences between 10-year-old boys and girls on their 3D mental rotation task. Our finding highlights the importance for future research to investigate sex differences in the extreme performance groups.

Furthermore, we did not find a sex difference on MRT-C performance on a group level in the older age group. This suggests that the task is too easy for use with older children (adolescents above 10 years of age). When a task is too easy, performance is subject to a ceiling effect. That is, all groups perform nearly perfectly within the time they have to perform the task, and so there is no scope for detection of group differences. This notion is substantiated by the fact that 44% of the youngest children belonged to the top 33% highest performers, whereas 67% of the oldest age group belonged to the 33% highest performers. Our finding that the mean performance of boys and girls is similar in the older group is also substantiated by additional analyses. These show that older boys and girls make an equivalent number of mistakes. This is an important finding and implies that age-appropriate tasks should be used when assessing cognitive abilities such as 3D mental rotation. New investigations of potential sex differences in MRT-C performance in children aged 9 years and older should be performed with a more difficult version of the task. The task difficulty can easily be increased by expanding the range of possible rotations (e.g., to between 0 and 360° around the vertical or horizontal axis, as in the VMRT; Neuburger et al., 2015)."

## Practical Implications

This study has a practical implication with regard to the stimulation of mental rotation skills and related spatial activities in children who lag behind in this function, notably young girls. It is known that spatial activities – such as spatial navigation and experiences in spatial play – could stimulate the maturation of brain networks underlying mental rotation ability (e.g., Krendl et al., 2008; Haier et al., 2009; Jaušovec and Jaušovec, 2012; Jolles and Crone, 2012; Dunst et al., 2013; Lowrie et al., 2017). Upon the structural changes in the brain, also the function of the brain areas involved improve (Jolles and Crone, 2012). For instance, Hawes et al. (2013) and Fernández-Méndez et al. (2018) showed that young children learn mental rotation as a result of carefully designed activities and lessons targeting the cognitive skill. Also, Nazareth et al. (2013) showed that the significant relation between the sex of the participant and MRT score is partially mediated by the number of masculine spatial activities participants had engaged in during their youth. Performing spatial activities thus both improves brain maturation and mental rotation skills. On this basis, it can be hypothesized that boys and girls develop similar mental rotation abilities when they are equally exposed to relevant spatial activities and encouraged to perform such activities (Nazareth et al., 2013).

There are various activities that involve spatial cognition, such as those involving the engagement of the total body while navigating throughout the environment. Other activities require more subtle motor movements, such as when building a tower out of wooden blocks. These kinds of activities require mental rotation of the wooden blocks and the to-be-build tower (Jansen and Heil, 2010). The importance of such activities to

school achievements has recently been provided by Giles et al. (2018). They showed that young children with better eye-to-hand coordination were more likely to achieve higher scores for reading, writing, and math. Teachers and caregivers should therefore encourage girls to engage in a variety of such spatial activities inside and outside of school. This is important because of the importance of spatial abilities for many daily life activities, such as finding one's way in three-dimensional space (e.g., to go to school, sports and playing games, see Newcombe and Frick, 2010). In addition, children's mental rotation abilities are fundamental to quantitative reasoning, such as in mathematics and geometrics, which requires the use of spatial cues, making comparisons and mentally visualizing, rotating and calculating the sides two- and three-dimensional figures (Nuttall et al., 2005; Rosselli et al., 2009; Jirout and Newcombe, 2015). Improving girls' mental rotation performance may lead to later success and achievement in the domain of STEM.

## CONCLUSION

This study showed that 3D mental rotation ability was positively correlated to mathematical achievement in 7–12-year-old children. We also showed that sex differences in 3D mental rotation emerge at least at the age of 7 years, to the advantage of boys. These findings are important with respect to improving sex differences in mathematical achievements and in STEM related disciplines. They suggest that interventions that stimulate the development of spatial skills may facilitate mathematical achievements, especially of young girls. Based on our results, we conclude that MRT-C is suitable for young children. Nevertheless, our results highlight the need to use age-appropriate tasks when assessing cognitive abilities, as we did not find sex differences in mean performance of children aged 10- to 12-years old. Future research is needed to fine-tune the MRT-C to make it suitable to both younger and older children.

## AUTHOR CONTRIBUTIONS

MvT gave substantial contributions to the conception and design of the work, analysis and interpretation of the data for the work, drafting of the work and final approval of the version to be published, and finally gave agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. MvdD and RdG gave substantial contributions to the design of the work, interpretation of the data for the work, critical revision of the work for important intellectual content and final approval of the version to be published, and finally gave agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JJ gave substantial contributions to the conception or design of the work, acquisition and interpretation of data for the work, drafting of the work, critical revision of the work for

important intellectual content and final approval of the version to be published, and finally agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## REFERENCES

- Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., and Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Dev. Neuropsychol.* 20, 385–406. doi: 10.1207/S15326942DN2001-5
- Boone, A. P., and Hegarty, M. (2017). Sex differences in mental rotation tasks: not just in the mental rotation process! *J. Exp. Psychol.* 43:1005. doi: 10.1037/xlm0000370
- Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. (1992). Remembering pictures: pleasure and arousal in memory. *J. Exp. Psychol.* 18, 379–390. doi: 10.1037/0278-7393.18.2.379
- Bruce, C. D., and Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: what is it? Why does it matter? And what can we do about it? *ZDM* 47, 331–343. doi: 10.1007/s11858-014-0637-4
- Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., and Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cogn. Instruct.* 26, 269–309. doi: 10.1080/07370000802177177
- Casey, B. M., Pezaris, E., Fineman, B., Pollock, A., Demers, L., and Dearing, E. (2015). A longitudinal analysis of early spatial skills compared to arithmetic and verbal skills as predictors of fifth-grade girls' math reasoning. *Learn. Individ. Diff.* 40, 90–100. doi: 10.1016/j.lindif.2015.03.028
- Central Office for Statistics (2016). *Proximity of Neighborhood Facilities: [Database: Nabijheid Voorzieningen Buurtcijfers 2016]*. New Delhi: Central Office for Statistics.
- Cherney, I. D., and London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5-to 13-year-old children. *Sex Roles* 54:717. doi: 10.1007/s11199-006-9037-8
- Christianson, S. A., and Loftus, E. (1990). Some characteristics of people's traumatic memories. *Bull. Psychon. Soc.* 28, 195–198. doi: 10.3758/bf03334001
- Dekker, S., Krabbendam, L., Aben, A., de Groot, R., and Jolles, J. (2013). Coding task performance in early adolescence: a large-scale controlled study into boy-girl differences. *Front. Psychol.* 4:550. doi: 10.3389/fpsyg.2013.00550
- Delgado, A. R., and Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence* 32, 25–32. doi: 10.1016/S0160-2896(03)00061-8
- Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Dunst, B., Benedek, M., Bergner, S., Athenstaedt, U., and Neubauer, A. C. (2013). Sex differences in neural efficiency: are they due to stereotype threat effect? *Pers. Individ. Diff.* 55, 744–749. doi: 10.1016/j.paid.2013.06.007
- Feng, J., Spence, I., and Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychol. Sci.* 18, 850–855. doi: 10.1111/j.1467-9280.2007.01990.x
- Fernández-Méndez, L. M., Contreras, M. J., and Elosúa, M. R. (2018). From what age is mental rotation training effective? Differences in preschool age but not in sex. *Front. Psychol.* 9:753. doi: 10.3389/fpsyg.2018.00753
- Frick, A. (2018). Spatial transformation abilities and their relation to later mathematics performance. *Psychol. Res.* doi: 10.1007/s00426-018-1008-5 [Epub ahead of print].
- Frick, A., Hansen, M. A., and Newcombe, N. S. (2013). Development of mental rotation in 3-to 5-year-old children. *Cogn. Dev.* 28, 386–399. doi: 10.1016/j.cogdev.2013.06.002
- Giles, O. T., Shire, K. A., Hill, L. J. B., Mushtaq, F., Waterman, A., Holt, R. J., et al. (2018). Hitting the target: mathematical attainment in children is related to interceptive-timing ability. *Psychol. Sci.* 29, 1334–1345. doi: 10.1177/0956797618772502
- Hahn, N., Heil, M., and Jansen, P. (2010). Preschoolers' mental rotation: sex differences in hemispheric asymmetry. *J. Cogn. Neurosci.* 22, 1244–1250. doi: 10.1162/jocn.2009.21236
- Haier, R. J., Karama, S., Leyba, L., and Jung, R. E. (2009). MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual-spatial task. *BMC Res. Notes* 2:174. doi: 10.1186/1756-0500-2-174
- Hango, D. W. (2013). *Gender Differences in Science, Technology, Engineering, Mathematics and Computer Science (STEM) Programs at University*. Ottawa: Statistics Canada.
- Hawes, Z., LeFevre, J. A., Xu, C., and Bruce, C. D. (2015). Mental rotation with tangible three-dimensional objects: a new measure sensitive to developmental differences in 4-to 8-year-old children. *Mind Brain Educ.* 9, 10–18. doi: 10.1111/mbe.12051
- Hawes, Z., Moss, J., Chang, D., and Naqvi, S. (2013). "Math for young children: a professional development model that bridges cognitive science and mathematics education," in *Proceedings of the 35th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, (Chicago, IL: University of Illinois at Chicago).
- Heil, M., and Jansen-Osmann, P. (2008). Gender differences in math and mental rotation accuracy but not in mental rotation speed in 8-years-old children. *Int. J. Dev. Sci.* 2, 190–196.
- Hoyek, N., Collet, C., Fargier, P., and Guillot, A. (2011). The use of the vanderberg and kuse mental rotation test in children. *J. Individ. Diff.* 33, 62–67. doi: 10.1027/1614-0001/a000063
- Jansen, P., and Heil, M. (2010). The relation between motor development and mental rotation ability in 5-to 6-year-old children. *Int. J. Dev. Sci.* 4, 67–75. doi: 10.3233/DEV-2010-4105
- Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., and Heil, M. (2013). Mental rotation performance in primary school age children: are there gender differences in chronometric tests? *Cogn. Dev.* 28, 51–62. doi: 10.1016/j.cogdev.2012.08.005
- Janssen, J., Verhelst, N., Engelen, R., and Scheltens, F. (2010). *Wetenschappelijke Verantwoording Papieren Toetsen Rekenen-Wiskunde Groep 3 tot en met 8 [Scientific Justification of the Mathematics Test]*. Arnhem: Cito.
- Jaušovec, N., and Jaušovec, K. (2012). Sex differences in mental rotation and cortical activation patterns: can training change them? *Intelligence* 40, 151–162. doi: 10.1016/j.intell.2012.01.005
- Jirout, J. J., and Newcombe, N. S. (2015). Building blocks for developing spatial skills: evidence from a large, representative US sample. *Psychol. Sci.* 26, 302–310. doi: 10.1177/0956797614563338
- Jolles, D. D., and Crone, E. A. (2012). Training the developing brain: a neurocognitive perspective. *Front. Hum. Neurosci.* 6:76. doi: 10.3389/fnhum.2012.00076
- Jolles, J. (2016). *Het Tienerbrein: Over de Adolescent Tussen Biologie en Omgeving*. Amsterdam: Amsterdam University Press.
- 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
- Kyttälä, M., and Lehto, J. E. (2008). Some factors underlying mathematical performance: the role of visuospatial working memory and non-verbal intelligence. *Eur. J. Psychol. Educ.* 23:77. doi: 10.1007/BF03173141
- Lowrie, T., Logan, T., and Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *Br. J. Educ. Psychol.* 87, 170–186. doi: 10.1111/bjep.12142
- Miller, D. I., and Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends Cogn. Sci.* 18, 37–45. doi: 10.1016/j.tics.2013.10.011

## ACKNOWLEDGMENTS

We gratefully thank Mathilde van Gerwen and Zita de Snoo for their assistance and their important contribution to the organization and execution of the BrainSquare study.



- Moè, A. (2016). Does experience with spatial school subjects favour girls' mental rotation performance? *Learn. Individ. Diff.* 47, 11–16. doi: 10.1016/j.lindif.2015.12.007
- Moè, A. (2018). Mental rotation and mathematics: gender-stereotyped beliefs and relationships in primary school children. *Learn. Individ. Diff.* 61, 172–180. doi: 10.1016/j.lindif.2017.12.002
- Moore, D. S., and Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy* 16, 435–445. doi: 10.1111/j.1532-7078.2010.00058.x
- Nazareth, A., Herrera, A., and Pruden, S. M. (2013). Explaining sex differences in mental rotation: role of spatial activity experience. *Cogn. Proc.* 14, 201–204. doi: 10.1007/s10339-013-0542-8
- Neuburger, S., Ruthsatz, V., Jansen, P., and Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learn. Individ. Diff.* 37, 169–175. doi: 10.1016/j.lindif.2014.09.003
- Newcombe, N. S., and Frick, A. (2010). Early education for spatial intelligence: why, what, and how. *Mind Brain Educ.* 4, 102–111. doi: 10.1111/j.1751-228X.2010.01089.x
- Nuttall, R. L., Casey, M. B., and Pezaris, E. (2005). *Spatial Ability as a Mediator of Gender Differences on Mathematics Tests: A Biological-Environmental Framework*. Cambridge: Cambridge University Press.
- Örnkloo, H., and von Hofsten, C. (2007). Fitting objects into holes: on the development of spatial cognition skills. *Dev. Psychol.* 43:404. doi: 10.1037/0012-1649.43.2.404
- Peters, M., and Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain Cogn.* 66, 260–264. doi: 10.1016/j.bandc.2007.09.003
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., and Richardson, C. (1995). A redrawn vanderberg and kuse mental rotations test-different versions and factors that affect performance. *Brain Cogn.* 28, 39–58. doi: 10.1006/brcg.1995.1032
- Rom, D. M. (2013). An improved hochberg procedure for multiple tests of significance. *Br. J. Math. Stat. Psychol.* 66, 189–196. doi: 10.1111/j.2044-8317.2012.02042.x
- Rosselli, M., Ardila, A., Matute, E., and Inozemtseva, O. (2009). Gender differences and cognitive correlates of mathematical skills in school-aged children. *Child Neuropsychol.* 15, 216–231. doi: 10.1080/09297040802195205
- Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703. doi: 10.1126/science.171.3972.701
- Status Scores (2016). *Central Office for Statistics [Database: Statusscores]*. Available at: [https://www.scp.nl/Onderzoek/Lopend\\_onderzoek/A\\_Z\\_alle\\_lopende\\_onderzoeken/Statusscores](https://www.scp.nl/Onderzoek/Lopend_onderzoek/A_Z_alle_lopende_onderzoeken/Statusscores)
- Thompson, J. M., Nuerk, H. C., Moeller, K., and Kadosh, R. C. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychol.* 144, 324–331. doi: 10.1016/j.actpsy.2013.05.009
- Titze, C., Jansen, P., and Heil, M. (2010). Mental rotation performance and the effect of gender in fourth graders and adults. *Eur. J. Dev. Psychol.* 7, 432–444. doi: 10.1080/17405620802548214
- Tolar, T. D., Lederberg, A. R., and Fletcher, J. M. (2009). A structural model of algebra achievement: computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educ. Psychol.* 29, 239–266. doi: 10.1080/01443410802708903
- Uttal, D. H., Miller, D. I., and Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: links to achievement in science, technology, engineering, and mathematics? *Curr. Dir. Psychol. Sci.* 22, 367–373. doi: 10.1177/0963721413484756
- van Tetering, M. A., and Jolles, J. (2017). Teacher evaluations of executive functioning in schoolchildren aged 9–12 and the influence of age, sex, level of parental education. *Front. Psychol.* 8:481. doi: 10.3389/fpsyg.2017.00481
- Vandenberg, S. G., and Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Mot. Skills* 47, 599–604. doi: 10.2466/pms.1978.47.2.599
- Voyer, D., Voyer, S., and Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* 117, 250–270. doi: 10.1037/0033-2909.117.2.250
- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127
- Wei, W., Yuan, H., Chen, C., and Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *Br. J. Educ. Psychol.* 82, 157–181. doi: 10.1111/j.2044-8279.2011.02049.x
- Zelazo, P. D., and Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: development and plasticity. *Child Dev. Perspect.* 6, 354–360. doi: 10.1111/j.1750-8606.2012.00246.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 © 2019 van Tetering, van der Donk, de Groot and Jolles. 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.





# Mental Number Representations in 2D Space

Elena Sixtus<sup>1,2\*</sup>, Jan Lonnemann<sup>2</sup>, Martin H. Fischer<sup>3</sup> and Karsten Werner<sup>1</sup>

<sup>1</sup> Faculty of Human Sciences: Research Group "Motor Control and Cognition," University of Potsdam, Potsdam, Germany,

<sup>2</sup> Empirical Childhood Research, University of Potsdam, Potsdam, Germany, <sup>3</sup> Division of Cognitive Sciences, University of Potsdam, Potsdam, Germany

## OPEN ACCESS

### Edited by:

Firat Soylu,  
The University of Alabama,  
United States

### Reviewed by:

Elizabeth Yael Toomarian,  
University of Wisconsin–Madison,  
United States  
Stephen Darling,  
Queen Margaret University,  
United Kingdom  
Ana Cristina Pires,  
Universidade de Lisboa, Portugal

### \*Correspondence:

Elena Sixtus  
esixtus@uni-potsdam.de

### Specialty section:

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

**Received:** 05 October 2018

**Accepted:** 18 January 2019

**Published:** 05 February 2019

### Citation:

Sixtus E, Lonnemann J,  
Fischer MH and Werner K (2019)  
Mental Number Representations  
in 2D Space. *Front. Psychol.* 10:172.  
doi: 10.3389/fpsyg.2019.00172

There is evidence both for mental number representations along a horizontal mental number line with larger numbers to the right of smaller numbers (for Western cultures) and a physically grounded, vertical representation where “more is up.” Few studies have compared effects in the horizontal and vertical dimension and none so far have combined both dimensions within a single paradigm where numerical magnitude was task-irrelevant and none of the dimensions was primed by a response dimension. We now investigated number representations over both dimensions, building on findings that mental representations of numbers and space co-activate each other. In a Go/No-go experiment, participants were auditorily primed with a relatively small or large number and then visually presented with quasi-randomly distributed distractor symbols and one Arabic target number (in Go trials only). Participants pressed a central button whenever they detected the target number and otherwise refrained from responding. Responses were not more efficient when small numbers were presented to the left and large numbers to the right. However, results indicated that large numbers were associated with upper space more strongly than small numbers. This suggests that in two-dimensional space when no response dimension is given, numbers are conceptually associated with vertical, but not horizontal space.

**Keywords:** spatial-numerical associations, SNARC, vertical space, horizontal space, Go/No-go task

## INTRODUCTION

The mental representation of numbers is a current and partly controversial subject. Especially studies regarding the SNARC (spatial-numerical association of response codes) effect are accumulating. Typically, such studies require participants to classify single digits presented at the center of a computer screen with speeded button responses as either odd or even. It is usually reported that in Western cultures, responses to numbers representing small magnitudes are faster on the left than right side and responses to numbers representing large magnitudes are faster on the right than left side (e.g., Dehaene et al., 1993; Schwarz and Keus, 2004; Keus and Schwarz, 2005; Wood et al., 2008; Hesse and Bremmer, 2017; Sixtus et al., 2017; Gökyaydin et al., 2018; Lohmann et al., 2018). Typically, the SNARC effect is explained as reflecting a horizontal mental number line (MNL) with larger numbers to the right of smaller numbers for Western cultures. Previous research suggests that this horizontal association partly depends on individual experiences. One example of such experiences are cultural conventions such as reading direction (Dehaene et al., 1993; Shaki and Fischer, 2008; Shaki et al., 2009; Fischer et al., 2010; see Nuerk et al., 2015 for a discussion

of mechanisms contributing to the influence of reading direction). Another relevant experience is finger counting, as indicated by an influence of the starting hand in finger counting on spatial-numerical associations (SNAs; Fischer, 2008; Fabbri, 2013; but see Sixtus et al., 2018). Moreover, there is evidence that there might be innate associations between magnitudes and horizontal space, both in humans and in animals (Rugani et al., 2015; de Hevia et al., 2017). In addition to this horizontal association of numbers and space, some studies investigated the diagonal, vertical, and radial relationship where usually numerically small numbers are associated with lower, lower left, and near space and numerically larger numbers with upper, upper right, and far space, respectively (e.g., Schwarz and Keus, 2004; Holmes and Lourenco, 2012; Sell and Kaschak, 2012; Fabbri, 2013; Grade et al., 2013; Winter and Matlock, 2013; Göbel, 2015; Winter et al., 2015; Hesse and Bremmer, 2017). Winter et al. (2015) discussed SNAs along three dimensions (horizontal, vertical, and radial) with an eye on potential origins of these different mappings and how these number mappings fit in with our current knowledge of brain organization and brain-culture interactions. Importantly, Fischer (2012; see also Fischer and Brugger, 2011; Pezzulo et al., 2011) suggested that the vertical association should be more stable than the horizontal one because it results from experience with laws of physics (e.g., a pile containing a larger number of objects is higher). Moreover, this vertical spatial association of magnitude is considered to be universal because physical laws apply regardless of cultural habits.

The present study aims to investigate conceptual spatial associations of numbers in two-dimensional space and in particular to compare relative association strengths along a horizontal and a vertical axis. “Conceptual” refers to the idea that spatial associations are an essential part of the meaning of numbers, rather than an extraneous and epiphenomenal part of number processing. In almost all studies, the way of assessing SNAs has not allowed for conclusions regarding conceptual SNAs because either the specific numerical magnitude or position in space (or both) have been explicit, task-relevant parts of experiments. Especially the association of numbers with horizontal space (i.e., spatial positions along a one-dimensional horizontal axis) has been investigated in a broad variety of tasks. In many paradigms, numbers are judged as regarding their parity (as odd or even) or magnitude (as smaller or larger than a reference number) and responses are given via buttons at the left and right side (ever since Dehaene et al., 1993). A magnitude judgment task with spatially distributed response buttons emphasizes both the magnitude represented by numbers and the spatial dimension along which numbers might be arranged. Thus, it addresses explicit magnitude processing and explicit spatial processing, which primes the spatial dimension along which response buttons are arranged within the task. In a study by Ranzini et al. (2016), participants were primed with a left- or rightward moving dot which they pursued with eye-movements and responded verbally (i.e., non-spatially) in a parity judgment task. Here, no spatial dimension was primed by responses and implicit magnitude processing was tested because parity judgments do not require magnitude information *per se*. However, the horizontal dimension was again clearly primed by

the horizontally moving dot. There are many more examples of studies which reported typical horizontal and/or vertical SNAs in very different paradigms but primed the spatial dimension, for example by head turns (horizontal: Loetscher et al., 2008; Sosson et al., 2018; horizontal and vertical: Winter and Matlock, 2013), left- or right turns when walking (Shaki and Fischer, 2014), and saccades (horizontal: Fischer et al., 2004; horizontal and vertical: Schwarz and Keus, 2004).

A recent study by Shaki and Fischer (2018) shed some light on conceptual spatial associations of numbers by testing both magnitude processing and spatial processing implicitly. More precisely, the authors compared explicit and implicit magnitude processing by employing both a magnitude comparison and a parity judgment task, respectively. They furthermore used an implicit association task (IAT), that is, a Go/No-go task with only one central response button so that a number-space association could not be primed by the response dimension but was determined by response rules alone. Target stimuli were single-digit numbers and arrows pointing left/right/down/up with the horizontal and vertical dimension in separate blocks. Response rules always combined number magnitude or parity with a directional cue (e.g., respond only to odd numbers and to arrows pointing leftward). The idea was that implicit associations between numbers and spatial concepts (arrows pointing toward different directions) influenced response efficiency. Indeed, the authors found that reliable horizontal associations failed to appear for implicit magnitude processing while vertical associations persisted (in their Experiment 1). This suggests that numbers are conceptually associated with vertical space only and that horizontal associations are merely an artifact of the task ingredients of usual SNARC experiments (e.g., priming horizontal space with spatially distributed response buttons). The present study follows up on Shaki and Fischer’s (2018) approach of addressing both implicit magnitude processing and implicit spatial processing.

However, even in Shaki and Fischer’s (2018) approach, one may argue that either the horizontal or vertical dimension were primed because each response rule included only one spatial dimension among the target items. The present study combines the horizontal and vertical dimension within one paradigm and compares relative association strengths along the two axes when both dimensions are relevant at the same time. Additionally to Shaki and Fischer’s (2018) study, so far only few studies have compared horizontal and vertical associations of numbers and those employed both dimensions within separate blocks or combined both in the form of diagonal axes. Such studies moreover led to differing interpretations. We will shortly describe four relevant studies in the following paragraphs.

Winter and Matlock (2013) employed a random number generation task where participants were instructed to randomly generate (and state) numbers while alternately turning the head to the left and right – or down and up in a separate block. The authors analyzed the average difference between a generated number and the one generated during the preceding head turn. Analogously to typical horizontal and vertical SNARC effects with faster left than right responses as well as faster down than up responses to small numbers (and vice versa for large numbers),

the authors expected smaller generated numbers after head turns toward to left side and downward, and larger generated numbers after head turns toward the right side and upward. Consistent with this hypothesis numbers generated during right- and upward head turns were on average significantly larger than the previously generated numbers. This effect was stronger in the vertical head turn condition than in the horizontal head turn condition, which suggests that vertical associations of numbers and space are stronger than horizontal associations. Blini et al. (2018) used optokinetic stimulation to induce shifts in spatial attention. Participants solved addition and subtraction problems during horizontal or vertical optokinetic stimulation. Besides specific effects of vertical optokinetic stimulation on decade errors during subtraction, gaze positions were influenced by operation type. Importantly, vertical eye movements were affected by operation type more reliably than horizontal eye movements, again suggesting that number processing (in this case addition and subtraction) interacts more strongly with vertical than horizontal spatial associations.

However, Holmes and Lourenco (2012) report contradictory results. The authors compared horizontal and vertical manual responses to numbers [0–9] in a parity judgment task as well as the two diagonal alignments of response positions. In their experiments, an Arabic digit appeared centrally on a touchscreen and responses were given via touches to visually presented response boxes below/above, left/right, left-below/right-above, or left-above/right-below the presented number. Importantly, the various response axes only appeared in separate blocks or experiments. A horizontal SNARC effect emerged with faster left than right responses to small numbers and faster right than left responses to large numbers, but no consistent vertical SNARC effect became evident. The diagonal SNARC depended on the horizontal axis, that is, SNAs apparently ran from left (down/up) for small numbers to right (up/down) for large numbers. Only in a second experiment when participants were instructed to imagine the numbers as floors in a building or levels of depth in a swimming pool did a vertical SNARC emerge. Hesse and Bremmer (2017) furthermore compared horizontal, vertical, and diagonal saccadic responses to numbers [1–9, except 5]. The parity status of numbers, which were presented visually and auditorily in separate blocks, had to be indicated via a saccade toward one of two dots that were positioned along a horizontal/vertical/diagonal axis. Again, only one axis was employed during an experimental block. The authors reported reliable SNARC effects with visually presented target numbers for the horizontal axis and the left-down to right-up diagonal axis. For the vertical axis, it was only present in error rates but not RTs and for the left-up to right-down diagonal axis it was not present at all.

Taken together, in the various studies, different pictures emerged depending on the task: it is unclear whether the vertical or horizontal representation of numbers is more reliable and, above all, it remains unclear whether the measured SNAs only arise as a result of the spatial presets of the concurrent task or response condition. Some indications of automatic co-activations among space and numerical magnitude are given by eye movement studies. Loetscher et al. (2010) found that

in a random number generation task changes in horizontal and vertical eye position correlated with number magnitude: right- and upward eye movements predicted the generation of a number larger than the previous one and left- and downward eye movements predicted the generation of a number smaller than the previous one. Hartmann et al. (2016) found that counting upward induced shifts of eye position up and to the right, while results were unclear for the task of counting downward. In a study by Holmes et al. (2016), participants were (digitally) dealt cards in a blackjack game which required the mental addition of the cards' numerical values. The authors found that the total numerical value of the dealt cards was reflected in the participants' eye movements along the horizontal axis. These studies suggest a conceptual spatial association of numerical operations (random number generation, counting, addition). However, it remains unclear whether specific numbers are also linked to spatial positions.

Another open question concerns the generalizability of specific SNAs. Galton (1880; see also Seron et al., 1992) presents descriptions of spatial arrangements of mental number representations which had been reported to him by various people. Many of these arrangements differ a lot from each other and also from the arrangement along a straight horizontal or vertical number line which is usually assumed in more recent studies. In number-form synesthetes, these spatial associations of numbers have even been shown to affect psychometric measures in spatial-numerical tasks (Jarick et al., 2009). This illustrates that mental representations may vary substantially inter-individually and that individual mental representations can have objective effects in measures of mental number processing. A study by Fischer and Campens (2009) also shows different spontaneous orientations of the number line when blindfolded participants were instructed to indicate spatial positions of different numbers by pointing somewhere in the space in front of them. Even in these non-synesthetic participants various orientations of the number lines emerged (horizontal, vertical, and radial). Note also that it is usual in SNARC experiments that the typical SNARC effect appears in only a part of the participants (typically between 60–80%; cf. Cipora and Wood, 2017; Wood et al., 2006a,b, 2008). Furthermore, there might be concrete factors which influence the specific orientation of individual MNLs (for a recent review, see Toomarian and Hubbard, 2018). This has been reported for reading direction (Shaki and Fischer, 2008; Shaki et al., 2009; Fischer et al., 2010) and the starting hand in finger counting (Fischer, 2008; Fabbri, 2013). Regarding reading direction, Shaki et al. (2009) found a reverse SNARC effect in Palestinians who read words and numbers from right to left. Regarding finger counting habits, Fischer (2008) found that participants who started to count on the left hand ("left-starters") but not those who started to count on the right hand ("right-starters") showed a reliable horizontal SNARC effect in a parity judgment task. On the other hand, in Fabbri's (2013) sample, right-starters showed a significantly stronger horizontal SNARC effect in a magnitude comparison task (but not in a parity judgment task). Although the results of these latter two studies do not fully converge, both suggest that experience with numbers through finger counting influences specific SNAs.

The main goal of the present study was to investigate the mental association between numbers and space in a two-dimensional grid of items when neither dimension (horizontal/vertical/diagonal) was primed by the task and when number magnitude was task-irrelevant. Avoiding a spatial distribution of response locations required a different method of integrating space into the task. In the present experiment, participants were therefore required to detect Arabic numbers within a grid of spatially distributed visual items and to respond with a single central response button. That way spatial congruency arose from the relationship between the numerical magnitude of the target number and the spatial position of visual target number presentation. Target numbers were auditorily primed before visual presentation and we expected that this auditory perception of numbers would co-activate associated spatial representations and would thereby influence search behavior and/or spatial attention. We included auditory primes that were numerically identical to the visual targets, so that number size could be expected to influence task performance before target detection. Thus, response efficiency at different spatial locations should be affected. Based on Holmes and Lourenco's (2012) and Hesse and Bremmer's (2017) findings, we would expect horizontal associations to be stronger than vertical associations. However, in these studies, spatially distributed responses were employed so that it is conceivable that the response dimension served as a prime for the number-space association (cf. Shaki and Fischer, 2018). Based on Shaki and Fischer's (2018) results, on the other hand, we would expect only a vertical, but no horizontal association. The present experiment therefore tests their argumentation that without explicit magnitude processing and/or horizontally arranged responses numbers should only be associated with vertical but not with horizontal space. We furthermore inquired about participants' finger counting habits to gain further insights into the relationship between individual finger-to-number mappings and SNAs.

## MATERIALS AND METHODS

### Participants

Thirty-seven participants were tested in return for payment. Data from two participants were excluded because of technical issues (causing partial data loss) or high error rate (see section "Analysis"). Five were non-German native speakers and were also excluded from analyses because the paradigm included German prime words. Of the remaining 30 participants, 23 were female and the mean age was 25 years,  $SD = 6.95$ . Handedness was assessed by self-report: one was left-handed, the rest were right-handers. The study was reviewed and approved by the Ethics Committee of the University of Potsdam, all subjects gave written informed consent, and the experiment was conducted in accordance with the ethical standards expressed in the Declaration of Helsinki.

### Apparatus

Participants were individually tested while seated at a table facing a PC monitor. Visual stimuli were presented on the PC monitor (60 Hz refresh rate, 68.5 cm screen diagonal). Auditory

stimuli were presented via headphones (AKG K-182; Harman Deutschland GmbH; Garching, Germany). The experiment was controlled and data recorded by experiment software (Krause and Lindemann, 2014) on a laptop (Lenovo T430s, Stuttgart, Germany). Participants were seated so that they had a viewing distance of 60 cm from the monitor. Midsagittally in front of the participants was a custom-made wooden box containing a central single response button (28 mm diameter).

### Stimuli

Primes were auditorily presented German number words (1: "eins," 2: "zwei," 8: "acht," 9: "neun") with a duration of 500 ms each, spoken by a female voice. Target numbers were visually presented Arabic numerals within a grid of distractor symbols ("##"; text size of target and distractor symbols = 28 pixels, sans serif font type). Each target screen included 49 black symbols on a white background which were arranged as follows. Centrally on the screen was a square of  $1100 \times 1100$  pixels which was again divided into  $7 \times 7$  equal squares. Each of these 49 mini-squares contained one symbol (distractor or target number). The exact position of the symbol within the square was randomly selected with the only constraint that it had a distance of at least 10 pixels from the mini-square's border to prevent overlap of the symbols. The borders of the mini-squares were not visible at any time.

### Procedure

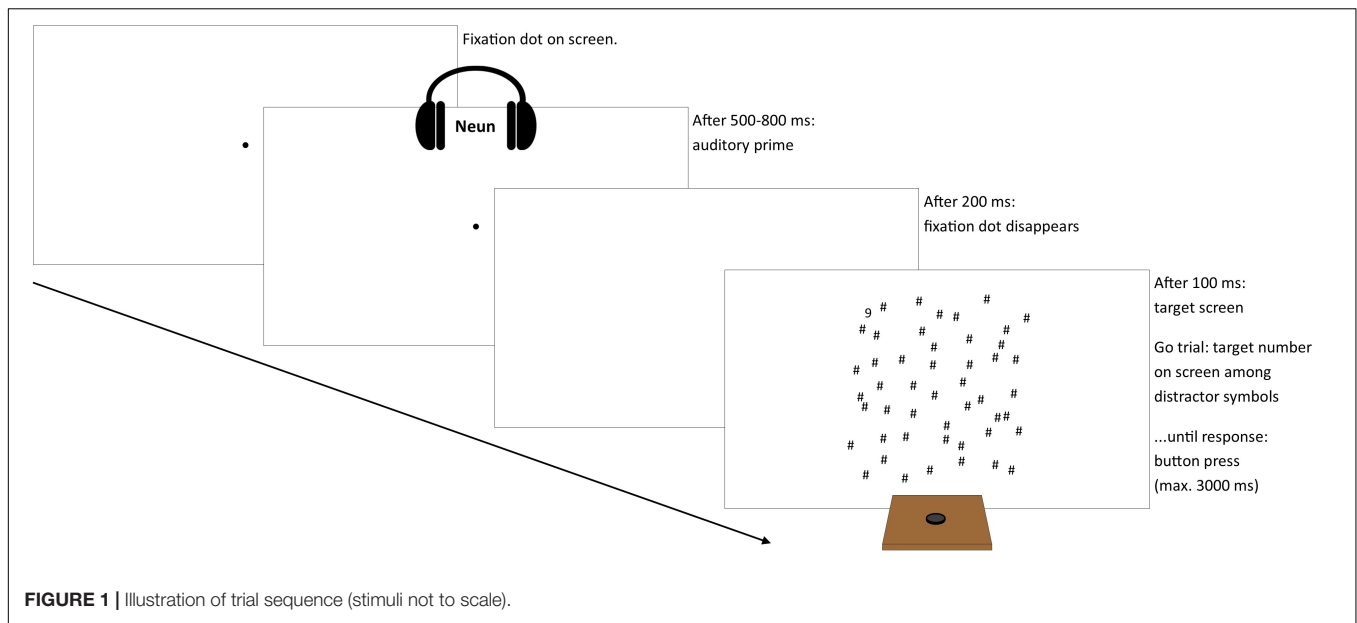
Participants were instructed to place both hands centrally in front of them with the dominant hand on the response button. Each trial started with the presentation of a central fixation dot. After a random interval between 500 and 800 ms, the auditory prime number was presented and 200 ms thereafter the fixation dot disappeared. After another 100 ms, the target screen appeared (see **Figure 1**). Participants responded by button press whenever they detected the Arabic numeral among the distractors (Go trials). Reaction times (RTs) were defined as the duration between the target screen presentation and button press. In No-go trials, there was no Arabic numeral on the screen and participants should refrain from responding. Each trial ended with a button press or after 3000 ms. Visual feedback was given after erroneous responses. Whenever participants made errors in two consecutive trials, a warning screen advised concentration. In the very beginning, there was a short training.

After the main experiment, the experimenter inquired the participant's spontaneous finger counting habits. She faced the participant, asked "Show me how you count from one to ten on your fingers," and noted the fingers used, the order of fingers, and the starting hand.

### Design

Each target number (1, 2, 8, 9) appeared at each of the 49 positions three times. Thus, the experiment comprised 588 Go trials ( $49 \text{ positions} \times 4 \text{ target numbers} \times 3 \text{ repetitions}$ ). Within each sequence of five Go trials, a No-go trial was inserted at a random position. On average, every 6th No-go trial was followed by an additional No-go trial to avoid predictability of Go trials following No-go trials. Auditory stimuli for the No-go trials were randomly selected among the target numbers.





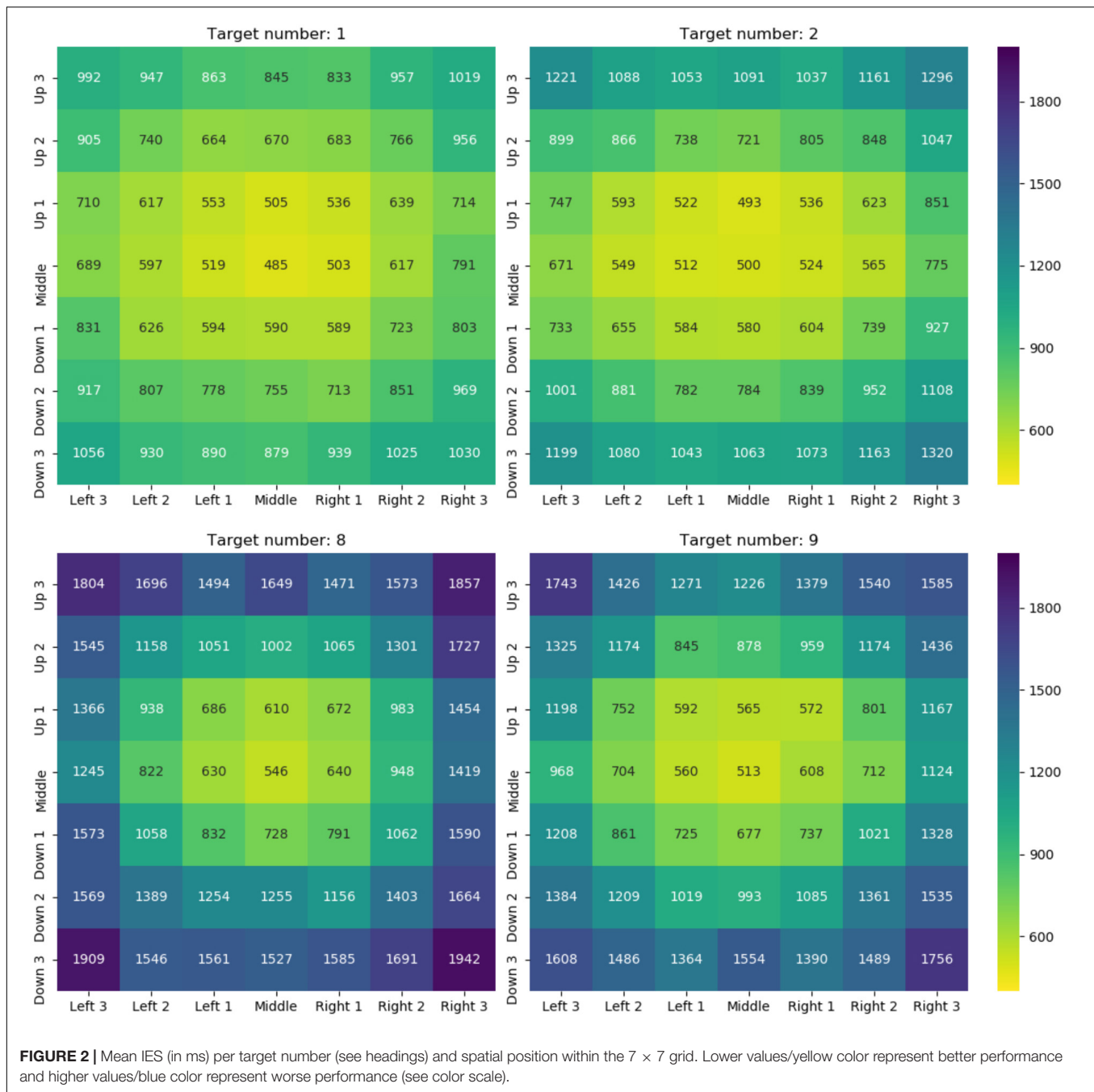
The training comprised 10 trials with at least four Go and four No-go trials. It could be repeated when necessary. The whole experiment took about 45 min.

## Analyses

Raw data and the analysis script are available online via <https://osf.io/4agk5>. Data of participants was excluded when the individual error rate in No-go trials exceeded the mean error rate in No-go trials among all participants plus/minus three standard deviations (SDs) because it indicates a tendency towards precocious responses without a genuine search of the target screen. This was the case for one participant with 25% erroneous No-go trials (mean error rate in No-go trials: 2.97%,  $SD = 4.92\%$  for  $N = 37$ ). After the exclusion of this participant, one participant with whom technical issues occurred, and the non-German native speakers (see section “Participants”) the mean error rate in Go trials was 1.58%,  $SD = 1.63\%$  and in No-go trials 2.38%,  $SD = 3.53\%$  ( $n = 30$ ). Trials with RTs below 300 ms were excluded (0.02% of the data). Only Go trials were further analyzed. Inverse efficiency scores (IES; e.g., Townsend, 1983; Bruyer and Brysbaert, 2013) are reported in the first place instead of raw RTs, because IES better reflect performance while RTs of correct responses neglect the worse performance of incorrect (i.e., missed) responses. IES was calculated as mean RTs of correct responses divided by the percentage of correct responses (PC) per participant for each condition relevant in the respective analysis (e.g., mean RTs/PC for left presentations and for right presentations). They are reported with ms as units and can be interpreted similar to RTs, that is, the smaller the IES, the more efficient (faster and accurate) was the response. An important precondition for using IES is the absence of a speed-accuracy trade-off. This was confirmed by preliminary analyses: RTs and error rates per target number and presentation position (i.e., specific horizontal and vertical position within the  $7 \times 7$  grid)

were strongly positively correlated,  $\rho = 0.74$ ,  $t(194) = 15.11$ ,  $p < 0.001$ .

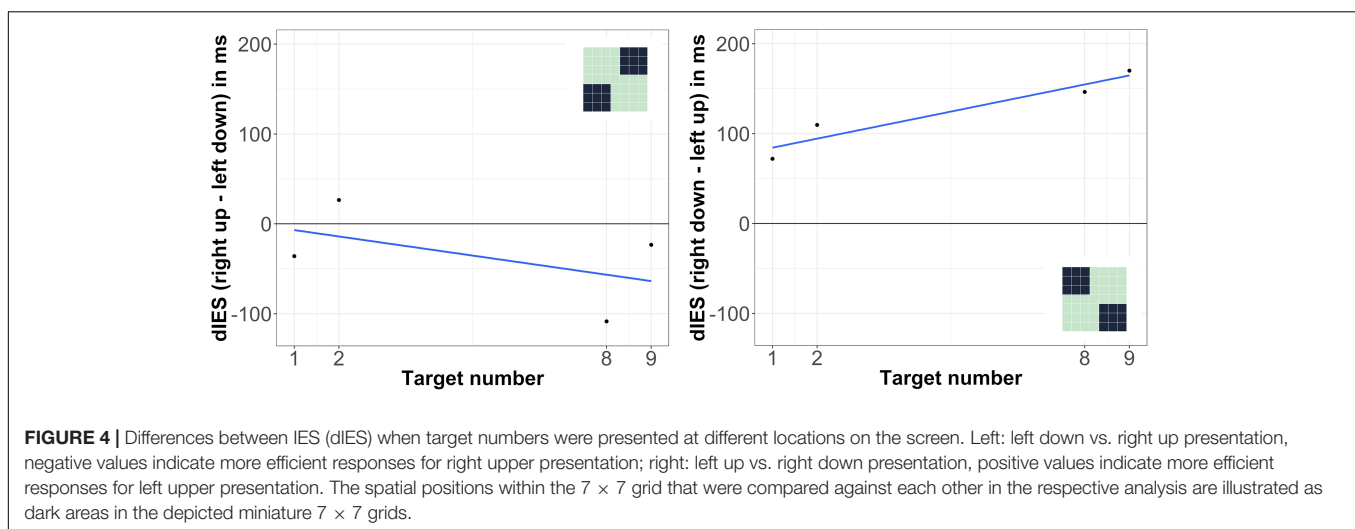
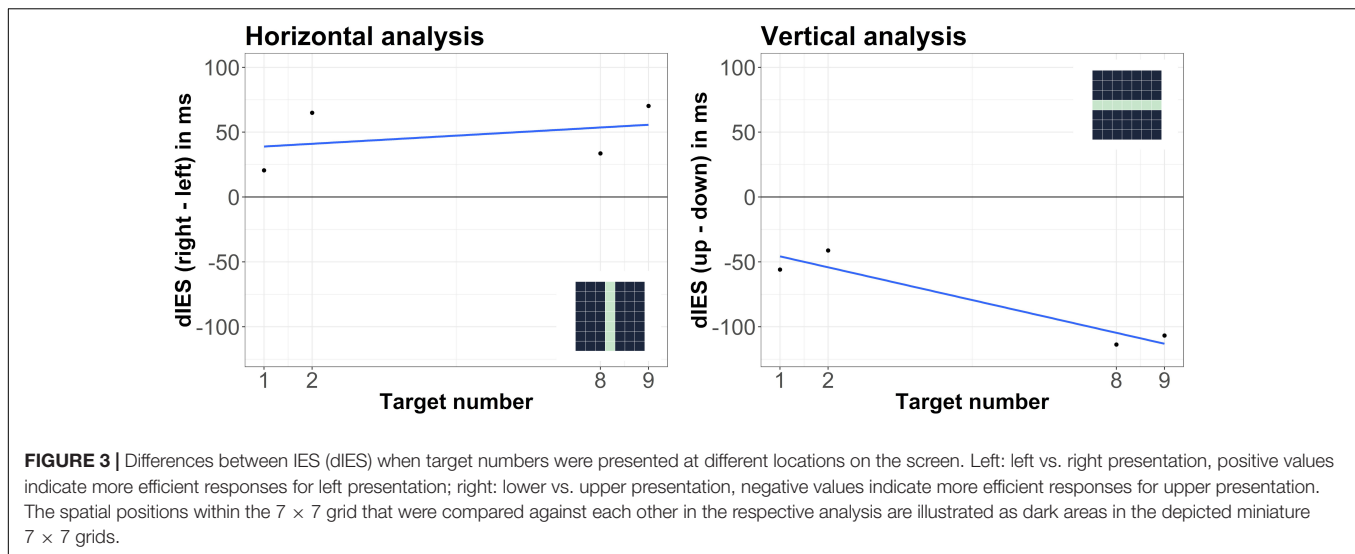
Being interested in SNAs similar to those reported in the literature on the SNARC effect, we aimed to follow the usual SNARC analyses. However, in contrast to usual SNARC experiments, the present experiment involved only one possible response with a central response button. Congruency arose from the relationship of numerical magnitude of the target number with the spatial position of visual target number presentation instead of the relationship of numerical magnitude of the target number with the spatial position of the response buttons. We analyzed both horizontal and vertical SNAs. The segmentation of the target screen into a  $7 \times 7$  grid allowed for a more or less central presentation of targets along the horizontal and vertical axis (i.e., targets within the seven, either horizontally or vertically aligned middle mini-squares). Analyses of responses to left-/right-/down-/upward targets therefore excluded trials with targets on the respective central positions. Potential effects of SNAs were analyzed separately for horizontal and vertical spatial segmentation (henceforth labeled “horizontal” and “vertical analysis”). Analogously to SNARC experiments, the individual IES differences (dIES) for responses to right-/upward minus left-/downward target presentations were calculated for each target (i.e., mean RTs/PC for right-/upward presentations minus mean RTs/PC for left-/downward presentations for each target). Thus, negative dIES values indicate more efficient responses to rightward presentation in horizontal analyses and to upward presentation in the vertical analyses. The individual dIES regression slopes over targets were tested against zero (e.g., Lorch and Myers, 1990; Pfister et al., 2013). Regression slopes from the horizontal and vertical analyses were compared against each other with a paired  $t$ -test. Effect sizes for  $t$ -tests were computed as Cohen’s  $d_z$  (cf. Lakens, 2013). All analyses were additionally conducted with RTs instead of IES for a more comprehensive assessment of the data.



For analyzing diagonal presentations, presentation positions on the target screen were segmented into four equal squares: left up, right up, left down, and right down (each including 3 × 3 mini-squares), excluding trials with targets on the respective central positions along the horizontal and vertical middle axis. Based on Hesse and Bremmer (2017), we calculated further dIES values: IES for right up-presentations minus IES for left down-presentations as well as IES for right down-presentations minus IES for left up-presentations per participant and target number. The resulting individual slopes over target numbers were again tested against zero.

## RESULTS

Preliminary analyses showed that in general IES (as well as RTs) increased with spatial distance of the target number from the center, all pairwise comparisons (Bonferroni-corrected) with  $p < 0.001$ , all  $d_z > 1.61$ . Furthermore, large target numbers (i.e., 8, 9) had larger IES (as well as RTs) than small target numbers (i.e., 1, 2). On average, IES for large numbers were larger than IES for small numbers by 379 ms,  $t(29) = 15.43$ ,  $p < 0.001$ ,  $d_z = 2.82$ . **Figure 2** depicts all mean IES per target number and spatial position.



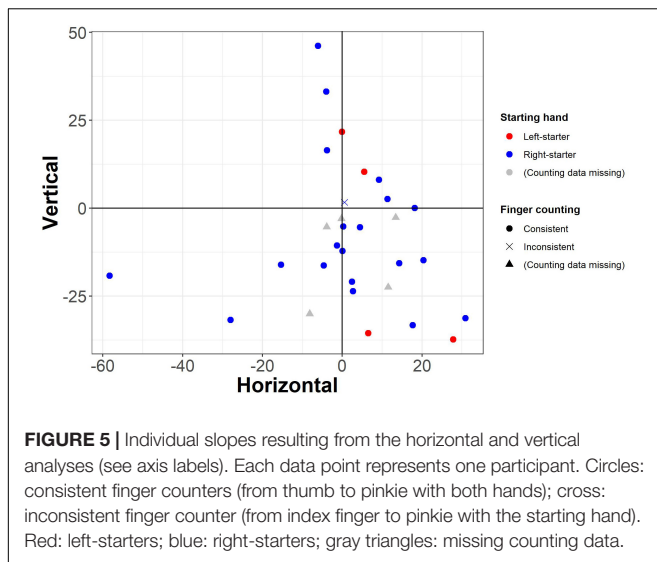
Horizontal and vertical analyses yielded a non-significant horizontal SNA and a significant vertical SNA: as visible in **Figure 3**, the slope in the horizontal analysis was not significantly different from zero,  $t(29) = 0.69$ ,  $p = 0.497$ , but the slope in the vertical analysis was,  $t(29) = -2.28$ ,  $p = 0.030$ ,  $d_z = 0.42$ . Moreover, the difference between the slopes was significant: the slopes from the vertical analysis were significantly larger (more negative) than the slopes from the horizontal analysis,  $t(29) = 2.10$ ,  $p = 0.045$ ,  $d_z = 0.38$ . Analyses of RTs (instead of IES) yielded similar results: the slope in the horizontal analysis was not significantly different from zero,  $t(29) = 0.42$ ,  $p = 0.677$ , but the slope in the vertical analyses was  $t(29) = -2.15$ ,  $p = 0.040$ ,  $d_z = 0.39$ . However, the difference between the two slopes was not significant,  $t(29) = 1.88$ ,  $p = 0.071$ . Overall, the results suggest that large numbers were more strongly associated with upper space than small numbers.

In the diagonal analyses, the on average negative slope for the analysis regarding the left down-right up axis was statistically not significant,  $t(29) = -1.32$ ,  $p = 0.196$ ; also the on average

positive slope for the analysis regarding the left up-right down axis was not significant,  $t(29) = 1.88$ ,  $p = 0.070$  (see **Figure 4**). Analyses of RTs (instead of IES) showed the same general – but non-significant – tendencies:  $t(29) = -1.27$ ,  $p = 0.213$  for the left down-right up axis;  $t(29) = 1.50$ ,  $p = 0.144$  for the left up-right down axis.

To explore individual differences in mental number representations, we furthermore compared individual slopes of the horizontal and vertical analyses and found that the two did not significantly correlate,  $\rho = -0.097$ ,  $t(28) = -0.52$ ,  $p = 0.610$ . As visible in **Figure 5**, the present data suggest SNAs following a down-small to up-large association for most participants: 21 of the 30 participants (i.e., 70%) had negative slopes in the vertical analysis which indicate more efficient responses to up- than downward presentations for larger numbers relative to smaller numbers.

Regarding finger counting habits, four participants were left-starters in finger counting, 21 were right-starters, and for five this information is missing. Twenty-nine counted from



thumb to pinkie with the starting and second hand for numbers 1–5 and 6–10, respectively (i.e., “consistent” counters). One participant counted from index finger to pinkie with the starting hand for numbers 1–4, used the full starting hand for number 5, and counted from thumb to pinkie of the second hand for numbers 6–10 (i.e., “inconsistent” counter). The effect of the starting hand and consistency of the fingers used in finger counting on horizontal and/or vertical SNAs was not statistically analyzed, because of the small sample size of left-starters ( $n = 4$ ) and inconsistent counters ( $n = 1$ ). Descriptively, the four left-starters did not share – and therefore did not point toward – a specific finger counting-dependent SNA pattern. Descriptive results are depicted in Figure 5.

## DISCUSSION

With the present study, we investigated mental representations of numbers in two-dimensional space. Importantly, our study extended previous research by combining the two spatial dimensions within a single paradigm without imposing the response location and thereby priming spatial congruency relations.

First of all, we compared SNAs in horizontal and vertical space and found that large numbers were more strongly associated with upper space than small numbers, implying a bottom-small to top-large directionality. Horizontal associations, however, neither significantly followed a left-to-right nor right-to-left directionality. When we focused on diagonally arranged presentation positions, results were not as clear-cut. Reflecting the results from the vertical analysis, large numbers were tendentially more strongly associated with upper left and upper right space (in comparison to lower right and lower left space, respectively) than small numbers, but the slopes did not significantly differ from zero. Note, however, that each of the diagonal analyses included only a subset of the

horizontal and vertical analyses. In an exploratory analysis, we furthermore compared individual SNAs for both horizontal and vertical associations and found no correlation between the two measures.

The dominance of the vertical association is in line with a recent study by Shaki and Fischer (2018) who argued that the horizontal SNARC effect “is an artifact of its measurement and number concepts are not inherently associated with horizontal space. The presence of horizontal SNAs (...) requires contextual priming” (p. 112). Regarding the vertical dimension, however, they provided “the first evidence for a purely conceptual SNA in this dimension” (p. 112). As in the present study, Shaki and Fischer’s (2018) experiment involved a Go/No-go task with only one central response button. Avoiding a spatial distribution of response buttons is essential to ensure that the spatial association under investigation is not created by the responses alone. The idea behind Shaki and Fischer’s (2018) study was that implicit associations between numbers and spatial concepts (arrows pointing toward different directions) influenced response efficiency. In the present experiment, the idea was that numbers influenced search behavior and/or spatial attention and thereby also response efficiency at different spatial locations. That is, the conceptualization of SNAs was not completely identical in that the spatial component in their case consisted of spatial concepts and in our case of spatial expectancies or attention shifts. Evidently, both kinds of SNAs exist with measurable effects in vertical space but not in horizontal space. The current results are also in line with the study by Blini et al. (2018) who found that mental arithmetic affected gaze positions. Here, SNAs refer to the relationship between gaze position and operation type. Vertical eye movements were more reliably affected (i.e., downward movements during subtractions and upward movements during additions) than horizontal eye movements. Taken together, evidence from a large variety of tasks is accumulating that vertical SNAs are more robust than horizontal SNAs.

On the other hand, the finding that the vertical association “trumped” the horizontal association seems to be in conflict with Holmes and Lourenco (2012; see also Hesse and Bremmer, 2017), where the horizontal association determined the SNARC slope more strongly than the vertical association when response buttons were arranged diagonally on the top left and bottom right. However, these divergent results are not surprising when taking into account Shaki and Fischer’s (2018) explanation that horizontal SNAs require contextual priming – which in this case consists in the presence of spatially arranged response buttons. Thus, horizontal SNAs can easily be primed and then may also be even stronger than vertical SNAs. Importantly, they fail to appear when their dimension is not primed while vertical SNAs persist. For a more general critique of the validity of diagonal SNAs, see Winter et al. (2015, p. 215).

Regarding the intra-individual comparison of horizontal and vertical SNAs, individual effects for horizontal and vertical associations did not seem to be related with one another: neither did participants have exclusive preferences for horizontal or vertical SNAs nor was there an “all-or-nothing” tendency with either both horizontal and vertical SNAs or none. Instead,



about two-thirds of participants exhibited vertical SNAs while only about one-third (partly overlapping) exhibited horizontal SNAs in the expected directions. Interestingly, the percentage of participants exhibiting vertical SNAs corresponds to the percentage of participants exhibiting horizontal SNAs in SNARC experiments in which the horizontal dimension is primed by task demands (between 60–80%; cf. Cipora and Wood, 2017; Wood et al., 2006a,b, 2008). A next step will be to determine whether participants who are responsive to horizontal primes are the same who resort to the vertical dimension in 2D space when no dimension is primed. In fact, although the present study extends previous research by combining the horizontal and vertical axis, it is still limited in its validity regarding individual conceptual SNAs. First of all, there is still one spatial dimension missing before comprehensive conclusions can be drawn regarding associations between numbers and all of space. Moreover, it only regards a limited range of numbers. As shown by Galton (1880), spatial associations of larger numbers can deviate substantially from those of single-digit numbers as used in the present experiment. Furthermore, mental space might not be as linearly arranged as the space employed by any kind of experiment. The best that can be done experimentally is to approach conceptual SNAs as closely as possible. As also shown by the present experiment, the exact orientation of SNAs seems to be a very idiosyncratic property with a more frequent occurrence of a preference for large numbers spatially above small numbers.

Furthermore, an as yet unmentioned finding of the present experiment was that responses were overall more efficient for target presentations in left and in upper space. A leftward bias might partly be explained by pseudoneglect, that is, an attentional bias toward left space in healthy persons (e.g., Jewell and McCourt, 2000). However, the fact that responses were also more efficient in upper space might suggest that visual search was affected by reading direction, which would be expected to begin at the upper left in 2D space. However, we were mainly interested in the relative efficiency of small and large numbers in 2D space, which is why we will not go into detail regarding general search behavior in our task.

The mechanism behind the reported SNA effect, that is, relatively faster RTs for large numbers in upper space, presumably involves an attentional shift and faster and/or preferred saccades toward the associated spatial position. Evidence comes from studies investigating horizontal SNAs for visually presented numbers. In a study by Fischer et al. (2003), visual attention was shifted toward the left or right side by mere visual perception of Arabic digits. While this seminal finding is

now under scrutiny (cf. Fischer and Knops, 2014), attentional consequences of number processing have now been extensively documented also in mental arithmetic (review in Fischer and Shaki, 2018). In addition, Fischer et al. (2004) investigated gaze durations after visual number presentations. Participants had to perform saccades to the left or right side of a centrally presented Arabic digit depending on the digit's parity. In responses to small numbers, leftward saccades were initiated faster than rightward saccades and vice versa for large numbers. Future studies employing spatially distributed target numbers in two-dimensional space could integrate eye tracking to further explore the impact of number magnitude on search behavior.

## CONCLUSION

The present study fills a gap as yet untouched by previous research: by arranging stimuli within a two-dimensional grid and thereby avoiding to prime any single axis, we extended studies on horizontal, vertical, and diagonal SNAs. Our main finding was that SNAs were predominantly determined by the vertical axis – with large numbers being more strongly associated with upper space than small numbers – while there was no specific preference for small vs. large numbers on the left vs. right side. Moreover, individual effects differed and we reported the relation of horizontal and vertical associations on an individual basis. Taken together, numbers seem to be conceptually associated with vertical but not horizontal space when number magnitude is task-irrelevant and neither spatial dimension is primed by task demands.

## AUTHOR CONTRIBUTIONS

ES, JL, and KW contributed to conception and design of the study. ES programmed the experiments, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to manuscript revision, and read and approved the submitted version.

## ACKNOWLEDGMENTS

We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam.

## REFERENCES

- Blini, E., Pitteri, M., and Zorzi, M. (2018). Spatial grounding of symbolic arithmetic: an investigation with optokinetic stimulation. *Psychol. Res. Adv.* doi: 10.1007/s00426-018-1053-0. [Epub ahead of print].
- Bruyer, R., and Brysbaert, M. (2013). Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychol. Belg.* 51:5. doi: 10.5334/pb-51-1-5
- 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 Hevia, M. D., Veggioni, L., Streri, A., and Bonn, C. D. (2017). At birth, humans associate “few” with left and “many” with right. *Curr. Biol.* 27, 3879.e2–3884.e2 doi: 10.1016/j.cub.2017.11.024
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371

- Fabbri, M. (2013). Finger counting habits and spatial-numerical association in horizontal and vertical orientations. *J. Cogn. Cult.* 13, 95–110. doi: 10.1163/15685373-12342086
- Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004
- Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cogn. Process.* 13(Suppl. 1), S161–S164. doi: 10.1007/s10339-012-0477-5
- Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260
- Fischer, M. H., and Campens, H. (2009). Pointing to numbers and grasping magnitudes. *Exp. Brain Res.* 192, 149–153. doi: 10.1007/s00221-008-1622-3
- Fischer, M. H., Castel, A. D., Dodd, M. D., and Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nat. Neurosci.* 6, 555–556. doi: 10.1038/nn1066
- Fischer, M. H., and Knops, A. (2014). Attentional cueing in numerical cognition. *Front. Psychol.* 5:1381. doi: 10.3389/fpsyg.2014.01381
- 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. (2018). Number concepts: abstract and embodied. *Philos. Trans. R. Soc. Lond. Series B Biol. Sci.* 373:125. doi: 10.1098/rstb.2017.0125
- Fischer, M. H., Warlop, N., Hill, R. L., and Fias, W. (2004). Oculomotor bias induced by number perception. *Exp. Psychol.* 51, 91–97. doi: 10.1027/1618-3169.51.2.91
- Galton, F. (1880). Visualized numerals. *Nature* 21, 252–256. doi: 10.1038/021252a0
- Göbel, S. M. (2015). Up or down? reading direction influences vertical counting direction in the horizontal plane - a cross-cultural comparison. *Front. Psychol.* 6:228. doi: 10.3389/fpsyg.2015.00228
- Gökaydin, D., Brugger, P., and Loetscher, T. (2018). Sequential effects in SNARC. *Sci. Rep.* 8:10996. doi: 10.1038/s41598-018-29337-2
- Grade, S., Lefevre, N., and Pesenti, M. (2013). Influence of gaze observation on random number generation. *Exp. Psychol.* 60, 122–130. doi: 10.1027/1618-3169/a000178
- Hartmann, M., Mast, F. W., and Fischer, M. H. (2016). Counting is a spatial process: evidence from eye movements. *Psychol. Res.* 80, 399–409. doi: 10.1007/s00426-015-0722-5
- Hesse, P. N., and Bremmer, F. (2017). The SNARC effect in two dimensions: evidence for a frontoparallel mental number plane. *Vis. Res.* 130, 85–96. doi: 10.1016/j.visres.2016.10.007
- Holmes, K. J., Ayzenberg, V., and Lourenco, S. F. (2016). Gamble on gaze: eye movements reflect the numerical value of blackjack hands. *Psychon. Bull. Rev.* 23, 1974–1981. doi: 10.3758/s13423-016-1055-0
- Holmes, K. J., and Lourenco, S. F. (2012). Orienting numbers in mental space: Horizontal organization trumps vertical. *Q. J. Exp. Psychol.* 65, 1044–1051. doi: 10.1080/17470218.2012.685079
- Jarick, M., Dixon, M. J., Maxwell, E. C., Nicholls, M. E. R., and Smilek, D. (2009). The ups and downs (and lefts and rights) of synaesthetic number forms: Validation from spatial cueing and SNARC-type tasks. *Cortex* 45, 1190–1199. doi: 10.1016/j.cortex.2009.04.015
- Jewell, G., and McCourt, M. E. (2000). Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia* 38, 93–110. doi: 10.1016/S0028-3932(99)00045-7
- Keus, I. M., and Schwarz, W. (2005). Searching for the functional locus of the SNARC effect: evidence for a response-related origin. *Mem. Cogn.* 33, 681–695. doi: 10.3758/BF03195335
- Krause, F., and Lindemann, O. (2014). Expyriment: a Python library for cognitive and neuroscientific experiments. *Behav. Res. Methods* 46, 416–428. doi: 10.3758/s13428-013-0390-6
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863
- Loetscher, T., Bockisch, C. J., Nicholls, M. E. R., and Brugger, P. (2010). Eye position predicts what number you have in mind. *Curr. Biol.* 20, R264–R265. doi: 10.1016/j.cub.2010.01.015
- Loetscher, T., Schwarz, U., Schubiger, M., and Brugger, P. (2008). Head turns bias the brain's internal random generator. *Curr. Biol.* 18, R60–R62. doi: 10.1016/j.cub.2007.11.015
- Lohmann, J., Schroeder, P. A., Nuerk, H.-C., Plewnia, C., and Butz, M. V. (2018). How deep is your SNARC? Interactions between numerical magnitude, response hands, and reachability in peripersonal space. *Front. Psychol.* 9:622. doi: 10.3389/fpsyg.2018.00622
- Lorch, R. F., and Myers, J. L. (1990). Regression analyses of repeated measures data in cognition research. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 149–157. doi: 10.1037/0278-7393.16.1.149
- Nuerk, H.-C., Patro, K., Cress, U., Schild, U., Friedrich, C. K., and Göbel, S. M. (2015). How space-number associations may be created in preliterate children: six distinct mechanisms. *Front. Psychol.* 6:215. doi: 10.3389/fpsyg.2015.00215
- Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. J. (2011). The mechanics of embodiment: a dialog on embodiment and computational modeling. *Front. Psychol.* 2:5. doi: 10.3389/fpsyg.2011.00005
- Pfister, R., Schroeder, P. A., and Kunde, W. (2013). SNARC struggles: instant control over spatial-numerical associations. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1953–1958. doi: 10.1037/a0032991
- Ranzini, M., Lisi, M., and Zorzi, M. (2016). Voluntary eye movements direct attention on the mental number space. *Psychol. Res.* 80, 389–398. doi: 10.1007/s00426-015-0741-2
- Rugani, R., Vallortigara, G., Priftis, K., and Regolin, L. (2015). Animal cognition Number-space mapping in the newborn chick resembles humans' mental number line. *Science* 347, 534–536. doi: 10.1126/science.aaa1379
- Schwarz, W., and Keus, I. M. (2004). Moving the eyes along the mental number line: comparing SNARC effects with saccadic and manual responses. *Percept. Psychophys.* 66, 651–664. doi: 10.3758/BF03194909
- Sell, A. J., and Kaschak, M. P. (2012). The comprehension of sentences involving quantity information affects responses on the up-down axis. *Psychon. Bull. Rev.* 19, 708–714. doi: 10.3758/s13423-012-0263-5
- Seron, X., Pesenti, M., Noël, M.-P., Deloche, G., and Cornet, J.-A. (1992). Images of numbers, or “when 98 is upper left and 6 sky blue”. *Cognition* 44, 159–196. doi: 10.1016/0010-0277(92)90053-K
- 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. (2014). Random walks on the mental number line. *Exp. Brain Res.* 232, 43–49. doi: 10.1007/s00221-013-3718-7
- Shaki, S., and Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition* 175, 109–113. doi: 10.1016/j.cognition.2018.02.022
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328
- Sixtus, E., Fischer, M. H., and Lindemann, O. (2017). Finger posing primes number comprehension. *Cogn. Process.* 3, 237–248. doi: 10.1007/s10339-017-0804-y
- Sixtus, E., Lindemann, O., and Fischer, M. H. (2018). Incidental counting: Speeded number naming through finger movements. *J. Cogn.* 1:44. doi: 10.5334/joc.49
- Sosson, C., Georges, C., Guillaume, M., Schuller, A.-M., and Schiltz, C. (2018). Developmental changes in the effect of active left and right head rotation on random number generation. *Front. Psychol.* 9:236. doi: 10.3389/fpsyg.2018.00236
- Toomarian, E. Y., and Hubbard, E. M. (2018). On the genesis of spatial-numerical associations: evolutionary and cultural factors co-construct the mental number line. *Neurosci. Biobehav. Rev.* 90, 184–199. doi: 10.1016/j.neubiorev.2018.04.010
- Townsend, J. T. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. New York, NY: Cambridge University Press.
- Winter, B., and Matlock, T. (2013). “More is up and right Random number generation along two axes,” in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 3789–3974.
- Winter, B., Matlock, T., Shaki, S., and Fischer, M. H. (2015). Mental number space in three dimensions. *Neurosci. Biobehav. Rev.* 57, 209–219. doi: 10.1016/j.neubiorev.2015.09.005

- 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. doi: 10.1016/S0010-9452(08)70219-3
- 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 © 2019 Sixtus, Lonnemann, Fischer and Werner. 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.



# Mathematics Competence Level: The Contribution of Non-symbolic and Spatial Magnitude Comparison Skills

Marisol Cueli<sup>1</sup>, Débora Areces<sup>1</sup>, Ursina McCaskey<sup>2,3</sup>, David Álvarez-García<sup>1</sup> and Paloma González-Castro<sup>1\*</sup>

<sup>1</sup> Department of Psychology, University of Oviedo, Oviedo, Spain, <sup>2</sup> Center for MR-Research, University Children's Hospital Zurich, Zurich, Switzerland, <sup>3</sup> Children's Research Center, University Children's Hospital Zurich, Zurich, Switzerland

## OPEN ACCESS

### Edited by:

Firat Soylu,  
The University of Alabama,  
United States

### Reviewed by:

Pom Charras,  
Paul Valéry University, Montpellier III,  
France  
Ana Miranda,  
University of Valencia, Spain

### \*Correspondence:

Paloma González-Castro  
mgcastro@uniovi.es

### Specialty section:

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

**Received:** 11 September 2018

**Accepted:** 15 February 2019

**Published:** 05 March 2019

### Citation:

Cueli M, Areces D, McCaskey U,  
Álvarez-García D and  
González-Castro P (2019)  
Mathematics Competence Level:  
The Contribution of Non-symbolic  
and Spatial Magnitude Comparison  
Skills. *Front. Psychol.* 10:465.  
doi: 10.3389/fpsyg.2019.00465

Magnitude comparison skills have been related to mathematics competence, although results in this area vary. The current study aimed to describe the performance of 75 children (aged 4–5 years) in two comparison tasks; and examine the strength of the relationship between each of the two tasks and mathematics competence level (MCL). Participants were assessed with the Early Numeracy Test which provides a global MCL score. Magnitude comparison skills were assessed with two tasks: a non-symbolic number comparison task and a spatial comparison task. Results of the Pearson correlation analysis showed a relationship between the two tasks with better performance in the spatial comparison task. Regression analysis with the stepwise method showed that only the non-symbolic number comparison task had a significant value in the prediction of the MCL pointing to the need to take these kinds of tasks into account in the first years of school.

**Keywords:** comparison skills, mathematics competence, non-symbolic comparison, preschool children, spatial comparison

## INTRODUCTION

A prominent characteristic of the majority of modern societies is the ubiquitous role of numeracy in conducting day-to-day activities (e.g., shopping or traveling requires the ability to make decisions based on quantitative information; Gilmore et al., 2013). Mathematical skills are therefore crucial abilities in modern life (Ancker and Kaufman, 2007) and early individual differences in mathematics have been reported to predict later adult socioeconomic status (Ritchie and Bates, 2013). Given this prominence, it is important to increase our knowledge of the cognitive processes underlying children's achievement in mathematics.

Findings from the Primary International Assessment Exercises which assess academic performance (International Association for the Evaluation of Educational Achievement [IEA], 2011; Organisation for Economic Co-operation and Development [OECD], 2014) warn about the existence of mathematical learning difficulties in children. Despite adequate and age-appropriate achievement in other educational domains, approximately 6–14% of school-age children have persistent difficulties with mathematics (Barbarelli et al., 2005; Clayton and Gilmore, 2015).

The study of cognitive determinants related to mathematical skills can be analyzed from either a domain general or a domain-specific perspective (Fias et al., 2013; Bellon et al., 2016). Domain general approaches focus on non-numerical cognitive skills that play a role in mathematical



performance, including executive functions such as working memory, processing speed, and inhibition control (Friso-van den Bos et al., 2013; Gilmore et al., 2013). Domain-specific approaches study the role of number-specific processes, such as individual differences in the representation of numerical magnitudes (Schneider et al., 2016). Domain-specific skills considered to be central to mathematics include procedural competence, conceptual understanding, counting, number fact knowledge, and Approximate Number System (ANS) acuity or “number sense” (Baroody, 2003).

## The Approximate Number System (ANS)

The ANS is a pre-linguistic cognitive system for representing and processing quantity information and has received a great deal of attention in recent years (Dehaene, 1997; Cantlon et al., 2009; Gallistel, 2011; Gilmore et al., 2014; Leibovich et al., 2017; Peng et al., 2017; Cai et al., 2018; Odic and Starr, 2018). The ANS supports the representation and processing of different magnitudes (Cantlon et al., 2009) and according to Dehaene (1997), it is a universal system present in animals, children, and adults.

Studies have shown that adults and children are able to use this system to compare and order sets of items presented as arrays of dots (Feigenson et al., 2004; Barth et al., 2005). The ANS allows comparison, addition, and subtraction of quantities without counting them (Dehaene, 1997). According to Odic et al. (2016), the ANS has three main characteristics: (1) Discrimination performance in the ANS is ratio-dependent based on Weber’s law (discriminating a collection of 12 items from six items is easier than discriminating a collection of 12 items from 11 items); (2) there are large individual differences in ANS precision (the ANS improves from birth until around age 30); and (3) the ANS has been located in both the human brain and in non-human animals (specifically, the ANS is associated with a region of the intraparietal sulcus).

An individual’s ANS acuity can be measured empirically with various tasks. These include symbolic (e.g., digit) or non-symbolic (e.g., dot) approximate comparison and addition tasks, or estimation tasks which assess the mapping between symbols and non-symbolic representations.

The most commonly used measure of ANS acuity is a dot comparison task, involving the comparison of two non-symbolic visual arrays of dots (Odic and Starr, 2018). During this task, participants see two dot arrays and must estimate which array they believe has more dots in it and respond either by key press, verbally, or by pointing. The response format used generally depends on the presentation methods employed and the age of the participants, and the performance is often indexed by accuracy (i.e., how often the participant correctly selects the more numerous array; Clayton and Gilmore, 2015). Performance on dot comparison tasks is affected by a distance effect and a size effect (Holloway and Ansari, 2010). The distance effect refers to the observation that decisions are more difficult when the numerical distance between the stimuli is small (in relation with the ratio-dependent effect). The size effect reflects more difficult discriminations for numerically larger numbers. This influence of task characteristics in how

children make decisions about number has been known since Piaget’s research, in which children erroneously judged one line of objects as more numerous when the objects were spaced further apart (Piaget, 1952). Starr et al. (2017) found that numerical decision-making in 4–6 year olds and adults was influenced by non-numerical features and when participants in their study were attempting to make decisions based on the numerosity of the arrays, even adults were unable to ignore the spacing of items within the arrays (although this effect decreased significantly with age).

The precision of the ANS in making non-symbolic comparisons improves with age (Halberda et al., 2008; Gómez-Velázquez et al., 2015; Odic et al., 2016), and has been proposed as a precursor of mathematical skills. Also, according to Odic and Starr (2018), the ANS is more precise in some people than in others and these individual differences emerge early in development and stay relatively stable with age (precision at 6 months predicts precision in preschool). Furthermore, individual differences in ANS precision demonstrate a small but significant relationship with formal math, including in preschoolers (Feigenson et al., 2013; Odic et al., 2016) and also correlate with the level of mathematics achievement (Halberda et al., 2008). For example, meta-analyses have reported significant correlations between ANS and mathematics (Chen and Li, 2014; Schneider et al., 2018) and many studies have shown a predictive association between number comparison skills and mathematical achievement (e.g., Piazza et al., 2010; Feigenson et al., 2013; Sasanguie et al., 2013; Starr et al., 2017). Starr et al. (2017) found that numerical acuity (measured with a non-symbolic comparison task) was the strongest predictor of variance in math achievement (although many other factors such as IQ and executive functions must be taken into account). Mazzocco et al. (2011) found that poor performance in non-symbolic approximation tasks distinguishes children with mathematical learning disabilities from their typically performing peers. Bonny and Lourenco (2012) found that preschoolers (3–5 years of age) with more precise number representations were generally more mathematically competent, as assessed by a standardized test of early math achievement. However, results in this area vary (Feigenson et al., 2013; Kroesbergen and Leseman, 2013) and not all studies have found significant links between non-symbolic number performance and mathematics achievement in children (e.g., Vanbinst et al., 2012; Kroesbergen and Leseman, 2013; Sasanguie et al., 2013). These differences in research findings could be related to the kind of tasks used to assess comparison skills.

One important issue that could affect performance in non-symbolic tasks is related to the dimension or representation of magnitude. Many researchers have suggested that number, time, and space are all represented by common mechanisms “a domain-general generalized magnitude system” (Odic et al., 2016). Walsh (2003) proposed a theory of magnitude (ATOM), which asserts that time, space, and number are all processed by this common magnitude system, located in parietal brain regions. Additional evidence for the generalized magnitude system comes from correlations of Weber fractions across

dimensions and from persistent congruency and interference effects between quantities, whereby manipulation of one dimension affects discrimination performance of another (Odic et al., 2016). However, authors such as Henik et al. (2017) suggested that the ability to perceive and evaluate sizes or amounts might constitute a more primitive system that has been used throughout human evolution as the basis for the development of the number sense and numerical abilities. How children and adults discriminate between different magnitudes has been analyzed from two perspectives: (1) the relationship between non-symbolic comparison, spatial comparison, and mathematics achievement and (2) the relationship between the performance in tasks with different magnitude systems (i.e., non-symbolic comparison and spatial comparison).

From the first perspective, Cai et al. (2018) examined the effects of symbolic (number line task) and non-symbolic estimation (point comparison task) on mathematics skills across three grade levels (kindergarten, children from grade 2, and children from grade 4). Their results showed that in kindergarten, non-symbolic estimation predicted all early mathematics skills while in grades 2 and 4, symbolic estimation accounted for unique variance in mathematical problem solving, but not in calculation fluency. Authors suggested that different types of ANS acuity should be used to predict mathematics skills in different learning periods and perhaps to identify children at risk of having difficulties in mathematics. Lourenco et al. (2012) tested the extent to which estimations of numerical and non-numerical magnitudes predicted math competence in college students. The tasks consisted of deciding which of two dot arrays was larger in either numerical value or cumulative surface area. Participants' accuracy scores on both magnitude tasks were positively correlated with performance on tests of advanced arithmetic. Later, Lourenco and Bonny (2017) used this procedure with 67 students between 5 and 6 years old who completed two magnitude comparison tasks (judge which of two discrete arrays was larger in numerical value and judge which of two amorphous displays was greater in cumulative area). They found that performance on number and area comparison tasks correlated with performance on exactly the same math tests and representations of cumulative area, and predicted children's math performance.

From the second perspective, Kucian et al. (2018) looked at the association between discrete non-symbolic number processing (comparison of dot arrays) and continuous spatial processing (comparison of angle sizes) in 367 children between the third and sixth grade. Their findings suggested that the processing of comparisons of dots and angles are related to each other, but angle processing was easier in their sample, so they concluded the existence of a more complex underlying magnitude system consisting of dissociated but closely interacting representations for continuous and discrete magnitudes. For this work, they used a task described in a previous study (McCaskey et al., 2017) which included a non-symbolic and a spatial task. Both tasks required a magnitude judgment, which is either based on discrete quantity

estimation of numerosity (number) or on continuous spatial processing (space). However, other authors such as Agrillo et al. (2013) did not find correlations between non-symbolic estimations (number/space/time) in 35 adults between 19 and 32 years old, which contradicts the existence of a general magnitude system.

Given the interest in the results of previous research (Agrillo et al., 2013; Cai et al., 2018; Kucian et al., 2018), and that Cai et al. (2018) found different performance in kindergarten and primary school, it would be useful to analyze the task set in McCaskey et al. (2017) in children younger than 6 years old and compare the results with those from Kucian et al. (2018) in older students between 8.2 and 12.9 years of age.

## The Present Study

The intention of the present study is to look deeply at the performance of preschool children when they have to make a magnitude judgment. Izard and Spelke (2009) showed that sensitivity detecting relationships of line length and angles improves over childhood, until 12 years of age. Furthermore, Starr et al. (2017) found that while 4-year-old children's numerical judgments were most influenced by non-numerical features, 6-year-old children exhibited strikingly adult-like performance, which suggested to these authors that numerical decision-making undergoes substantial change between 4 and 6 years of age.

With this in mind, this study aims to: (1) describe the performance of preschool children (aged 4–5 years) in the two magnitude comparison tasks used by McCaskey et al. (2017) and Kucian et al. (2018) and (2) examine the strength of the relationship between each of the two tasks (non-symbolic and spatial magnitude comparison) and mathematics competence level (MCL).

## MATERIALS AND METHODS

### Participants

Participants in this study were 75 students enrolled in three second-year kindergarten classes, in the Principality of Asturias (North of Spain). Schools were public and were located in a city-center. By law, classes must have no more than 25 students per class. All the families reported a medium-high socio-economic level and consisted of three to four members.

The students were aged between 4 and 5 years old ( $M = 52.47$ ,  $SD = 3.91$  months; in a range of 46–59 months). Of these students, 44 (59%) were girls and 31 (41%) were boys. There were no statistical differences in the gender-distribution of boys and girls in the current sample,  $\chi^2(1) = 2.25$ ,  $p = 0.13$ . Furthermore, differences in the MCL were not significant in terms of age ( $p = 0.228$ ), gender ( $p = 0.836$ ), or intelligence quotient (IQ;  $p = 0.275$ ) according to univariate analysis of variance. A convenience sample was recruited for the study. Written informed consent was obtained from the parents of the participants of this study. No children had been diagnosed with learning disabilities and all of them had an IQ between 80 and 130 ( $IQ\ M = 99.52$ ;  $SD = 14.99$ ).

## Measures

### Raven's Colored Progressive Matrices

Raven's progressive matrices provide a non-verbal assessment of intelligence. The test offers three progressively more difficult forms intended for different populations. Items on all forms ask the examinee to identify the missing component in a series of figurative patterns. In this study, the colored form (CPM; Raven et al., 1996) was used. This is used to assess children from 4 years of age. It consists of 36 items in three sets of 12. The administration time is usually 15–30 min.

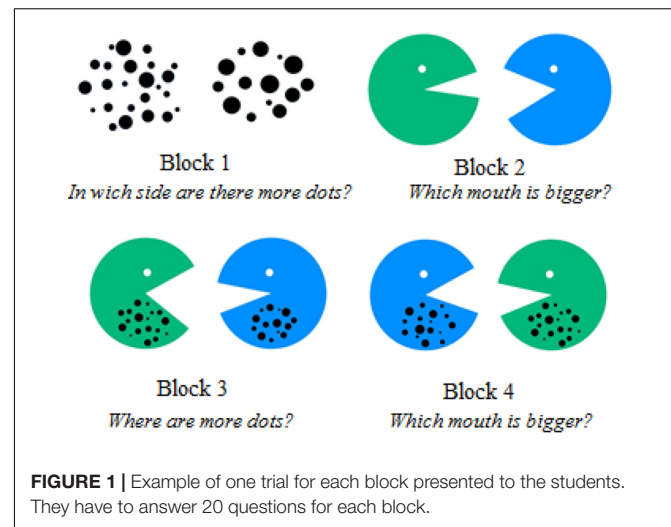
### Early Numeracy Test Revised

The original revision of the Early Numeracy Test – Revised (ENT-R; Navarro et al., 2009; Mendizábal et al., 2017) was completed by Van Luit and Van de Rijt (2009) and subsequently standardized for the Spanish population (Van Luit et al., 2015). The ENT-R evaluates early numerical knowledge and aims to detect students with mathematical learning disabilities. This tool is especially useful in the transition from preschool to elementary education. It can be used to confirm which students need support to cope with the new mathematical learning, thereby promoting the implementation of early intervention procedures. The test assesses eight skills: concepts of comparison, classification, one to one correspondence, seriation, verbal counting, structured counting, counting (without pointing), general knowledge of numbers, and estimation. A global MCL score can be obtained based on performance across the eight subtests. The ENT-R is suitable for children aged 4–7 years. There are three parallel versions of 45 items each, version A was used in the current study. It takes an average of 30 min to complete the test, which is individually administrated. Previous studies have reported a Cronbach's alpha reliability index of 0.95 (Mendizábal et al., 2017). In the current study, the Cronbach's alpha was 0.76. Only the global MCL score was used for analyzing the relationship with the two magnitude comparison tasks.

### Magnitude Comparison Tasks

Comparison skills were assessed by the test developed by McCaskey et al. (2017). It is based on two tasks: a non-symbolic number comparison task and a spatial comparison task. Both tasks require a magnitude judgment, which is either based on the evaluation of discrete quantity estimation of numerosity (number) or on continuous spatial processing (space). The first task is based on the presentation of two sets of dots. Children have to indicate on which side more black dots are presented. The second task shows a green and a blue pacman facing each other with varying mouth sizes. Children have to indicate which of the two pacman figures has a bigger mouth. The first is a non-symbolic number comparison task, and the second requires a visuo-spatial and continuous magnitude decision.

The tasks were presented using E-prime software (Version 2.0). There were 80 different trials, classified into four blocks (20 trial per block). In the first block “dots” (B1), in each trial two groups of dots ranging from a minimum of 12 to a maximum of 30 dots, were presented horizontally. Children were asked to indicate on which side more black dots were presented (Figure 1). Presentation of dots was controlled for individual size of dots (no



**FIGURE 1** | Example of one trial for each block presented to the students. They have to answer 20 questions for each block.

judgment possible due to individual dot size), total displayed area (no judgment possible due to total black area), distribution of dots (no judgment possible due to total covered area), and the numerical distance between presented magnitudes. All children were carefully introduced to the task and encouraged to solve all trials by comparison of both sets of presented dots by numerical estimation and highlighting the importance of not counting. Responding was indicated by pressing a key corresponding to the side of the larger magnitude (z key or m key).

In the second block “mouths” (B2), a green and a blue pacman facing each other with varying mouth sizes were presented horizontally. Children had to indicate which of the two pacman figures had a bigger mouth (Figure 1). In contrast to the non-symbolic number (dot) comparison task, this task required a visuo-spatial and continuous magnitude decision. The mouth angle of the pacman figures varied between a minimum of 27° to a maximum of 68°. The side of the correct answer and color of correct pacman were balanced. In the same way as for the number comparison task, children were carefully instructed and advised to solve the spatial comparison task by simple estimation of mouth sizes and not to use other approaches (e.g., their fingers, or any other tool) to measure the mouth sizes.

For the third and fourth blocks, the stimuli were combined and each presentation consisted of a green and a blue pacman facing each other with the dots presented inside the figures (Figure 1). In the third block “dots combined” (B3), the child was required to decide in which of the two sets there were more dots. In the fourth block, the child had to indicate which mouth size was bigger. In the third block, nine trials were congruent (more dots and a bigger mouth) and 11 trials were incongruent (fewer dots and a bigger mouth). Similarly, in the fourth block “mouths combined” (B4), there were nine congruent (a bigger mouth and more dots) and 11 incongruent trials (bigger mouth and fewer dots). The same stimuli were used for blocks 3 and 4, although order of presentation was randomized. Children were explicitly instructed to look at the dots (block 3) or at the mouths (block 4).

All tasks were administered in an untimed format following the procedures from other authors such as Defever et al. (2013)



and Szűcs et al. (2013). An untimed task allowed us to complete all the trials and avoid omissions in performance. Finally, the ratio between smaller and larger dot arrays and between smaller and larger mouth angle across all blocks was 0.5, 0.6, 0.7, 0.8, and 0.9. The same ratios were used in the four blocks but the order of presentation was randomized. In summary, the first and third blocks were based on a non-symbolic number comparison task and the second and fourth blocks were spatial comparison tasks. Prior to the start of each block, students performed four training trials with ratios of 0.5 and 0.6 to ensure that they understood the instructions. All students did the same training trials and received feedback during this initial practice.

## Procedure

After obtaining research approval (the study was approved by Ministry of Science, Innovation and Universities of Spain and by the University of Oviedo, Asturias, Spain), local preschools were randomly selected and approached to take part in the study. The schools forwarded the information about the study to parents of the children with a request for informed consent. The IQ of the children whose parents agreed to participate was assessed with the Raven's CPM. All children scored an IQ between 80 and 130 and were therefore included in the study, undergoing further testing with the ENT-R and the comparison task. All the assessment tasks were administered by qualified educational psychologists and were coordinated and guided by the same educational psychologist from the research group.

The study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013). The evaluations were carried out over consecutive days during regular classes.

## Data Analysis

Preliminary examination of the data showed that the assumptions (e.g., skewness and kurtosis) required for the use of parametric statistics were met. All analyses were conducted using SPSS for Windows Version 22. Differences were considered significant at level of  $p < 0.05$ . For both the non-symbolic number and the spatial comparison tasks, the accuracy or correct responses (CRs) were taken into account (CRs over total items).

Initially, to describe the performance on the two comparison tasks, Pearson correlation coefficients were calculated. In order to study this relationship in depth, the CR in every block was compared by paired student  $t$ -tests and effect sizes were calculated. For the interpretation of the effect sizes, Cohen (1988) criterion was used, which establishes that the effect is small when  $\eta^2 = 0.01$  ( $d = 0.20$ ), medium when  $\eta^2 = 0.059$  ( $d = 0.50$ ), and high if  $\eta^2 = 0.138$  ( $d = 0.80$ ).

Second, to examine the strength of the relationship between each of the two tasks and MCL, a hierarchical multiple regression analysis was carried out. We tested three models. The MCL was included in the analysis as the dependent variable. In the first model, gender, age, and IQ were used as independent variables; in the second model, the CR in block 1 (B1) and block 3 (B3) were added as independent variables (given that these blocks are based on a non-symbolic number comparison task); and in the third model, the CR in block 2 (B2) and block 4 (B4) were

taken also included (given that these blocks are based on a spatial comparison task).

## RESULTS

### Pearson Correlation Coefficients

The correlations are provided in **Table 1**, including the mean, standard deviation, skewness, and kurtosis of the four blocks (dots, mouth size, dots combined, and mouths combined).

As can be seen in **Table 1**, significant correlations were found between B1–B2, B1–B3 and between B2–B3, B2–B4, B3–B4. Also B1 and B3 showed a significant relationship with the MCL. In **Table 2**, the percentages of CRs are provided.

The  $t$ -test showed significant differences between B1–B2  $t(74) = -12.76$ ,  $p < 0.001$ ,  $d = 2.1$ ; B1–B4  $t(74) = -11.50$ ,  $p < 0.001$ ,  $d = 1.89$ ; B2–B3  $t(74) = 13.29$ ,  $p < 0.001$ ,  $d = 2.18$ ; B3–B4  $t(74) = -12.85$ ,  $p < 0.001$ ,  $d = 2.11$ . Differences were not significant between B1–B3 ( $p = 0.910$ ) and B2–B4 ( $p = 0.312$ ). Student performance was similar in the comparison of dots (B1) and dots inside the mouth (B3) and also in the performance in the comparison of mouths (B2) and mouths with dots inside (B4) which makes sense given the common

**TABLE 1 |** Correlation matrix of the magnitude comparison skills and MCL including means, standard deviations, skewness, and kurtosis.

	1. Block1	2. Block2	3. Block3	4. Block4	5. MCL
1. Block 1	–	0.24*	0.34**	0.09	0.40***
2. Block 2		–	0.31**	0.37**	0.21
3. Block 3			–	0.28*	0.27**
4. Block 4				–	–0.16
5. MCL					–
<i>M</i>	13.07	17.20	13.11	17.47	22.05
<i>SD</i>	2.65	1.74	2.63	2.22	14.87
Skewness	0.05	–0.54	0.04	–1.97	1.24
Kurtosis	–0.63	0.02	–0.44	5.78	2.19
Minimum	8	13	7	8	1
Maximum	19	20	19	20	76

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Note. *M* = mean; *SD* = standard deviation; efficacy = correct responses – incorrect responses; MCL = mathematics competence level provided by ENT-R; Block1 = comparing dots; block 2 = comparing mouth sizes; block 3 = comparing dots showed inside pacman mouths; block 4 = comparing pacman mouth sizes with dots inside.

**TABLE 2 |** Proportion of correct responses for the four blocks of the comparison task across the five ratio conditions.

Comparison task	Ratio				
	0.5	0.6	0.7	0.8	0.9
Block 1	68%	44%	21%	15%	5%
Block 2	93%	73%	72%	64%	13%
Block 3	77%	65%	51%	51%	3%
Block 4	81%	89%	69%	66%	53%

Note. Block 1 = comparing dots; block 2 = comparing mouths; block 3 = comparing dots combined; block 4 = comparing mouths combined.



nature of the tasks. The means indicate more CR in the execution of B2 and B4 showing more CRs when children have to compare mouths than when they have to compare dots. Furthermore, independently of the presentation of dots alone or inside the pacman mouth, the performance was similar with better results in B3 (dots inside the pacman) which could be related to the congruence effect. Similarly, the performance for the presentation of the pacman mouth alone or with dots inside did not differ.

## Regression Analysis

In a second step, we carried out hierarchical multiple regression analyses in order to analyze which of the comparison skills better predicts the MCL.

The MCL was taken as dependent variable and the CR of the four blocks as independent variables. In a first model, gender and age were used as independent variables. In the second model, age, gender, and CR of blocks 1 and 3 were included as independent variables. These two blocks assess non-symbolic comparison skills. Finally, in the third model, the CR of blocks 2 and 4 were introduced in addition to gender, age, CR of blocks 1 and 3. These two blocks assess spatial comparison skills. Results showed that the three models were significant with  $F(3,71) = 3.299$ ,  $p = 0.025$ ;  $F(5,69) = 3.928$ ,  $p = 0.003$ , and  $F(7,67) = 4.060$ ,  $p = 0.001$ , respectively. In the first model, IQ was significant ( $p = 0.032$ ). In the second model, B1 was significant ( $p = 0.019$ ) and in the third model B1 ( $p = 0.024$ ) and B4 ( $p = 0.010$ ) were significant.

Looking at  $R^2$  and the adjusted  $R^2$  for the three models, 12% of the variation in MCL can be explained by model 1, 22% by model 2, and 29% of the variation in the MCL can be explained by model 3 (see Table 3).

## DISCUSSION

The aim of this study was to analyze the performance of 4 year old children in two magnitude comparison tasks, a non-symbolic and a spatial comparison task. Furthermore, we were

interested in examining which of the two tasks was more strongly related with MCL.

Results showed that both tasks were significantly, positively related when dots or mouths were shown alone (B1 and B2) or combined (B3 and B4). The correlation between the blocks was 0.24 and 0.28. Kucian et al. (2018) obtained a very similar result with a correlation of 0.26 between the two tasks and 0.25 controlling for the effect of age and grade level. These authors were also interested in whether the strength of this correlation decreased with development. They observed that in fifth grade, the correlation was weaker than in lower grade levels, but they did not observe differences in sixth grade regarding previous levels. If we compare the 4-year-old students (in this study) with the third to sixth grade students (in Kucian et al., 2018) the correlations are very similar and *a priori*, they would not yield significant differences. In short, if we consider present and past results we can see that both tasks are significantly related to each other.

However, at the same time, we found differences in the performance in every task. Regarding the mean performance of the students, it seems that the non-symbolic comparison task was more difficult than the spatial comparison task. Students made more mistakes when comparing dots alone or inside the mouths. This result is in line with research from McCaskey et al. (2017) who showed a significant relationship between the two tasks and pointed out that when the ratios were similar (as in the present study), spatial judgment of angle size is easier compared to non-symbolic magnitude comparison. Leibovich and Henik (2013) pointed to higher accuracy levels for a continuous spatial task compared to non-symbolic dot comparison. Odic et al. (2013) showed higher acuity for continuous spatial processing (comparison of area sizes) than non-symbolic number processing (comparison of dot arrays) in 3–6-year-old children. Also, Kucian et al. (2018) found that in third to sixth grade students, spatial comparison was generally easier than non-symbolic number comparison. They showed significant differences between the number and spatial tasks in third, fourth, fifth, and sixth grades.

These differences in both tasks are also reflected in the performance in the four blocks with respect to the ratios, which

**TABLE 3 |** Hierarchical regression analysis models.

	Model 1		Model 2		Model 3	
	$\beta$ ( $\beta$ standardized)	$t$	$\beta$ ( $\beta$ standardized)	$t$	$\beta$ ( $\beta$ standardized)	$t$
Age	0.73(0.19)	1.71	0.48(0.12)	1.16	0.36(0.09)	0.89
Gender	−3.36(−0.11)	−1.00	−2.01(−0.06)	−0.61	−1.65(−0.05)	−0.52
IQ	0.26(0.24)	2.18*	0.16(0.15)	1.37	0.16(0.15)	1.40
B1			1.58(0.28)*	2.40*	1.47(0.26)*	2.31*
B3			0.65(0.11)	1.00	0.94(0.16)	1.43
B2					1.24(0.14)	1.22
B4					−1200(−0.30)*	−2.66*
$R^2$	0.122		0.222		0.298	
$\Delta R^2$	0.085		0.165		0.224	

\* $p < 0.05$ . Note. Values in the table are the non-standardized  $\beta$  regression coefficient; those in brackets are the standardized values.  $t$  = student  $t$ -test;  $R^2$  = variance explained;  $\Delta R^2$  = change in variance explained; B1 = correct responses of block 1 comparing dots; B2 = correct responses of block 2 comparing mouth; B3 = correct responses of block 3 comparing dots combined; B4 = correct responses of block 4 comparing mouths combined.

were the same in the four blocks (0.5, 0.6, 0.7, 0.8, and 0.9). In this case, we saw performance decrease as the ratio approached 1 and in consequence, the level of difficulty increased. The decrease is more evident in the case of the non-symbolic comparison task than in the spatial comparison task. Odic et al. (2016) highlighted that discrimination performance in the ANS is ratio-dependent based on Weber's law (the accuracy of this system varies as the quantity increases, with the comparison being easier for very different sets such as 10 versus 5). These results are also in line with previous work by Kucian et al. (2018), in which the accuracy levels decreased significantly for both conditions (non-symbolic magnitude comparison task and spatial comparison task) with increasing ratio between magnitudes (bigger ratios mean smaller distances between magnitudes and are therefore more difficult to compare). The authors hypothesized, similarly to Leibovich and Henik (2013), that the superiority of processing continuous magnitudes might indicate that this system is older than the system for processing discrete magnitudes and might develop earlier during childhood than the discrete quantity system. Our results point in the same direction and support this idea of previous and older development of the continuous magnitude system, although more research is still needed. Our second aim was to examine the strength of the relationship between each of the two tasks (non-symbolic and spatial magnitude comparison) and the level of mathematics competence (MCL). The correlation analysis showed that there was a relationship between the non-symbolic comparison and the MCL showing that the child's performance in this type of task is related to their level in mathematics but not performance in the spatial task in this sample. This has an immediate educational implication. When teachers analyze the performance of their students in comparison activities, it is very important for them to take into account performance in non-symbolic comparison tasks because that could be more related to their mathematics level in this age range. In this sense, the results in the dots comparison task are compatible with the findings of Mazzocco et al. (2011) or Feigenson et al. (2013) who highlighted that poor performance on non-symbolic approximation tasks distinguishes children with mathematical learning disabilities from their typically performing peers. The results from the spatial comparison task, which did not show a correlation with the MCL, differ from previous research (Lourenco et al., 2012; Cai et al., 2018). This could be related to the difficulty level given that although the two tasks in the study were comparable in their design (same ratios for the non-symbolic and spatial comparison task), that did not mean the same level of difficulty for 4-year-old children, and the second task was easier for them so it is possible that it did not discriminate sufficiently. This could be the reason for the absence of a relationship with the MCL reflected in the regression analysis.

The regression analyses showed that the dots comparison alone had a significant value in the prediction of the MCL (model 2). The mouths combined comparison was also significant in the explanation of the MCL (model 3). However, the dots comparison seems to have more weight in this prediction given that the value in B1 (dots comparison) was positive while the value in B4 (mouths combined comparison) was negative. In any

case, this supports the results found by Feigenson et al. (2013) and Sasanguie et al. (2013) showing a predictive association between number comparison and mathematical achievement. The mouths combined comparison exhibited a relationship with MCL, albeit negative. This result could be associated with the characteristics of the task. The second and fourth blocks included congruent and incongruent trials. The total number of trials in each block was 20, and this may not have been sufficient for accurate assessment when the two situations are included (nine congruent and 11 incongruent). The negative result is quite surprising, and needs to be replicated in the future, in order to understand whether the reason is associated with the congruent and incongruent trials, or whether children could be looking at other characteristics of the stimuli, or even whether this task is especially difficult for 4-year-old students. Children can answer by looking at the dots instead of the angles in the fourth block and for this reason, they may answer incorrectly in the incongruent trials (when the bigger angle has fewer dots). Starr et al. (2017) highlighted that performance is typically better for congruent trials compared to incongruent trials and the effect of congruency is strongest for young children and attenuates with age, suggesting that younger children may be more biased by non-numerical cues than older children. The influence of the congruence and incongruence effects could be the reason for the children's better performance in block 3 compared to block 1. In this sense, it is possible that tasks including congruence and incongruence for students between 4 and 5 years old are not appropriate to their level and do not provide significant information. However, this needs to be studied more in the future.

Taking the results together, we can see that comparison at 4 years old can be influenced by different aspects (magnitude used, congruency, characteristics of the stimuli as the density or size) that make it harder to interpret the children's performance. Non-symbolic comparison tasks (such as the dots comparison) may be more useful with simple designs including ratios lower than 0.7, given that ratios of 0.9 are extremely complicated for children at this educational level. However, in spatial comparison tasks (such as the mouths comparison), lower ratios are especially easy for children and the design of the tasks have to include ratios higher than 0.7 to improve discriminatory power. In addition, the use of congruency and incongruency has to be studied more deeply and could be analyzed in relation with the MCL. It is important to note that in this study we used congruency and incongruency in terms of their relationship between the mouths and the dots (more dots and a bigger mouth or fewer dots and a bigger mouth). Typically, congruency has been studied in terms of the features of the stimuli and considering trials congruent trials when one or more visual cues (dots area, density) are positively correlated with numerosity, and incongruent trials, when one or more visual cues are negatively correlated with numerosity. Several studies have demonstrated that these visual cues can influence numerosity judgments such as Gebuis et al. (2009) and it has been associated with other factors such as inhibition (Szűcs et al., 2013), so it could be interesting to analyze the profile of performance in the task in relation to the executive function levels.

Finally, this study has the following limitations that must be taken into account. First, sample selection by accessibility is a limitation of the study, although it is necessary given the difficulty of going into the schools and working with children as it affects the running of the school. Also, it is necessary to note that the sample size is rather small for multiple regression analysis but it allows us to draw preliminary conclusions in this line of research using this specific comparison task. However, given that the aim of the study was to determine the strength of the relationship between the MCL and the numerical and spatial tasks, and given the differences between the two tasks and the MCL, it would be useful to check these results in students of these ages and even to use more trials for each block of tasks to avoid possible ceiling effects. In addition, the MCL was taken as a global measure rather than using specific mathematic skills (classification, seriation, one to one correspondence, verbal counting...), it could be interesting in the future to examine the relationship of each specific mathematical skill to the two comparison tasks. In any case, in conclusion, the results of our study have a practical implication for teachers, showing that tasks associated with the comparison of dots could provide an approximate measure of students' MCL. At the same time, activities that require that comparison can enhance and improve students' MCL, so it might be interesting to incorporate these kinds of tasks in the objectives and instructional procedures for teaching mathematics in preschool. In short, even from the first years, teachers can have

an approximation of a student's MCL and improve it directly or indirectly through tasks of magnitude judgment such as the comparison of dots.

## AUTHOR CONTRIBUTIONS

MC, DA, and PG-C contributed to the design of this study. UM contributed to the design of the comparison task. MC and PG-C organized the data collection and database. DÁ-G performed the statistical analyses. MC and DA wrote the first draft of the manuscript. UM and PG-C wrote a major revision of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This work is funded by the I + D + i project with reference EDU2015-65023-P and by the regional project with reference FC-GRUPIN-IDI/2018/000199.

## ACKNOWLEDGMENTS

The authors wish to thank Nigel V. Marsh Ph.D. for his helpful comments on the manuscript.

## REFERENCES

- Agrillo, C., Piffer, L., and Adriano, A. (2013). Individual differences in non-symbolic numerical abilities predict mathematical achievements but contradict ATOM. *Behav. Brain Funct.* 9:26. doi: 10.1186/1744-9081-9-26
- Ancker, J. S., and Kaufman, D. (2007). Rethinking health numeracy: a multidisciplinary literature review. *J. Am. Med. Inform. Assoc.* 14, 713–721. doi: 10.1197/jamia.M2464
- Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., and Jacobsen, S. J. (2005). Math learning disorder: incidence in a population-based birth cohort, 1976-1982, Rochester Minn. *Ambul. Pediatr.* 5, 281–289. doi: 10.1367/A04-209R.1
- 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 Studies*, eds A. J. Baroody and A. Dowker (Mahwah: Lawrence Erlbaum Associates Inc), 1–33.
- 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
- Bellon, E., Fias, W., and De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval in children related to inhibition? *Front. Psychol.* 7:825. doi: 10.3389/fpsyg.2016.00825
- Bonny, J. W., and Lourenco, S. F. (2012). 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
- Cai, D., Zhang, L., Li, Y., Wei, W., and Georgiu, G. K. (2018). The role of approximate number system in different mathematics skills across grades. *Front. Psychol.* 9:1733. doi: 10.3389/fpsyg.2018.01733
- Cantlon, J. F., Platt, M. L., and Brannon, E. M. (2009). Beyond the number domain. *Trends Cogn. Sci.* 13, 83–91. doi: 10.1016/j.tics.2008.11.007
- Chen, Q., and Li, J. (2014). Association between individual differences in nonsymbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Clayton, S., and Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM* 47, 759–770. doi: 10.1007/s11858-014-0655-2
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd Edn. New Jersey, NJ: Lawrence Erlbaum.
- 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
- Dehaene, S. (1997). *The Number Sense*. Oxford: Oxford University Press.
- 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
- Feigenson, L., Libertus, M. E., and Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Dev. Perspect.* 7, 74–79. doi: 10.1111/cdep.12019
- Fias, W., Menon, V., and Szucs, D. (2013). Multiple components of development dyscalculia. *Trends Neurosci. Educ.* 2, 43–47. doi: 10.1016/j.tine.2013.06.006
- 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
- Gallistel, C. (2011). Prelinguistic thought. *Lang. Learn. Dev.* 7, 253–262. doi: 10.1080/15475441.2011.578548
- Gebuis, T., Cohen Kadosh, R., De Haan, E., and Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cogn. Process.* 10, 133–142. doi: 10.1007/s10339-008-0219-x
- 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
- Gilmore, C., Attridge, N., De Smedt, B., and Inglis, M. (2014). Measuring the approximate number system in children: exploring the relationships among different tasks. *Learn. Individ. Differ.* 29, 50–58. doi: 10.1016/j.lindif.2013.10.004
- Gómez-Velázquez, F. R., Berumena, G., and González-Carrido, A. A. (2015). Comparisons of numerical magnitudes in children with different levels of

- mathematical achievement. An ERP study. *Brain Res.* 1267, 189–200. doi: 10.1016/j.brainres.2015.09.009
- Halberda, J., Mazocco, 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
- Henik, A., Gliksman, Y., Kallai, A., and Leibovich, T. (2017). Size perception and the foundation of numerical processing. *Curr. Dir. Psychol.* 26, 45–51. doi: 10.1177/0963721416671323
- Holloway, I. D., and Ansari, D. (2010). Developmental specialization in the right intraparietal sulcus for the abstract representation of numerical magnitude. *J. Cogn. Neurosci.* 22, 2627–2637. doi: 10.1162/jocn.2009.21399
- International Association for the Evaluation of Educational Achievement [IEA] (2011). *Results of the 2011 PIRLS and TIMSS tests in Spain*. Madrid: Ministerio de Educación, Cultura y Deporte.
- Izard, V., and Spelke, E. S. (2009). *Development of sensitivity to geometry in visual forms*. *Hum. Evol.* 23, 213–248.
- 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: 10.3389/fpsyg.2018.01733
- Kucian, K., McCaskey, U., von Aster, M., and O’Gorman Tuura, R. (2018). Development of a possible general magnitude system for number and space. *Front. Psychol.* 9:2221. doi: 10.3389/fpsyg.2018.02221
- Leibovich, T., Al-Rubaiey Kadhim, S., and Ansari, D. (2017). Beyond comparison: the influence of physical size on number estimation is modulated by notation, range and spatial arrangement. *Acta Psychol.* 175, 33–41. doi: 10.1016/j.actpsy.2017.02.004
- Leibovich, T., and Henik, A. (2013). Comparing performance in discrete and continuous comparison tasks. *Q. J. Exp. Psychol.* 67, 899–917. doi: 10.1080/17470218.2013.837940
- Lourenco, S. F., and Bonny, J. W. (2017). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Dev. Sci.* 20:e12418. doi: 10.1111/desc.12418
- Lourenco, S. F., Bonny, J. W., Fernandez, E. P., and Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18737–18742. doi: 10.1073/pnas.1207212109
- Mazocco, 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
- McCaskey, U., von Aster, M., O’Gorman-Tuura, R., and Kucian, K. (2017). Adolescents with developmental dyscalculia do not have a generalized magnitude deficit – processing of discrete and continuous magnitudes. *Front. Hum. Neurosci.* 11:102. doi: 10.3389/fnhum.2017.00102
- Mendizábal, E. A., Villagrán, M. A., and Guzmán, J. I. N. (2017). Instructional support system for teaching number sense. *Rev. Educ.* 375, 14–33. doi: 10.4438/1988-592X-RE-2016-375-333
- Navarro, J. I., Aguilar, M., Alcalde, C., Marchena, E., Ruiz, G., Menacho, I., et al. (2009). Estimación del aprendizaje matemático mediante la versión española del test de evaluación matemática temprana de Utrecht [Assessment of math learning by the Early Mathematical Competence Test Spanish version]. *Eur. J. Educ. Psychol.* 2, 131–143. doi: 10.1989/ejep.v2i2.24
- Odic, D., Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Dev. Psychol.* 49, 1103–1112. doi: 10.1037/a0029472
- Odic, D., and Starr, A. (2018). An introduction to the approximate number system. *Child Dev. Perspect.* 12, 223–229. doi: 10.1111/cdep.12288
- Odic, D., Vale-Lisboa, J., Eisinger, R., González-Olivera, M., Maiche, A., and Halberda, J. (2016). Approximate number and approximate time discrimination each correlate with school math abilities in young children. *Acta Psychol.* 163, 17–26. doi: 10.1016/j.actpsy.2015.10.010
- Organisation for Economic Co-operation and Development [OECD] (2014). *PISA 2012. Programa para la evaluación internacional de los alumnos. Informe español resultados y contexto* [Programme for International Student Assessment. Spanish report]. Madrid: Ministerio de Educación, Cultura y Deporte.
- Peng, P., Yang, X., and Meng, X. (2017). The relation between approximate number system and early arithmetic: the mediation role of numerical knowledge. *J. Exp. Child Psychol.* 157, 111–124. doi: 10.1016/j.jecp.2016.12.011
- Piaget, J. (1952). *The Child’s Concept of Number*. New York, NY: Norton.
- 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
- Raven, J. C., Court, J. H., and Raven, J. (1996). *Raven Matrices Progresivas*. Madrid: TEA.
- Ritchie, S. J., and Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* 24, 1301–1308. doi: 10.1177/0956797612466268
- 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
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.1237
- Schneider, M., Merz, J., Sticker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., et al. (2018). Associations of number line estimation with mathematical competence: a meta-analysis. *Child Dev.* 89, 1467–1484. doi: 10.1111/cdev.13068
- Starr, A., DeWind, N. K., and Brannon, E. M. (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
- 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
- Van Luit, J. E. H., and Van de Rijt, B. A. M. (2009). *The Early Numeracy Test Revised*. Doetinchem: Graviant.
- Van Luit, J. E. H., Van de Rijt, B. A. M., Navarro, J. I., Aguilar, M., Alcalde, C., Marchena, E., et al. (2015). *Test de Evaluación de la Competencia Matemática Temprana-Informatizado (TEMT-i)* [The utrecht early mathematical competence test]. Madrid: EOS.
- Vanbinst, K., Ghesquière, P., and De Smedt, B. (2012). Numerical magnitude representations and individual differences in children’s arithmetic strategy use. *Mind Brain Educ.* 6, 129–136. doi: 10.1111/j.1751-228X.2012.anish report]. Madr01148.x
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002
- World Medical Association (2013). Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

**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 Cueli, Areces, McCaskey, Álvarez-García and González-Castro. 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.





# The Importance of Ordinal Information in Interpreting Number/Letter Line Data

Christine Podwysocki, Robert A. Reeve and Jason D. Forte\*

Melbourne School of Psychological Sciences, University of Melbourne, Parkville, VIC, Australia

## OPEN ACCESS

### Edited by:

Sharlene D. Newman,  
Indiana University Bloomington,  
United States

### Reviewed by:

Koleen McCrink,  
Columbia University, United States  
Carmelo Mario Vicario,  
University of Messina, Italy

### \*Correspondence:

Jason D. Forte  
jdf@unimelb.edu.au

### Specialty section:

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

**Received:** 28 September 2018

**Accepted:** 12 March 2019

**Published:** 27 March 2019

### Citation:

Podwysocki C, Reeve RA and  
Forte JD (2019) The Importance  
of Ordinal Information in Interpreting  
Number/Letter Line Data.  
Front. Psychol. 10:692.  
doi: 10.3389/fpsyg.2019.00692

The degree to which the ability to mark the location of numbers on a number-to-position (NP) task reflects a mental number line (MNL) representation, or a representation that supports ordered lists more generally, is yet to be resolved. Some argue that findings from linear equation modeling, often used to characterize NP task judgments, support the MNL hypothesis. Others claim that NP task judgments reflect strategic processes; while others suggest the MNL proposition could be extended to include ordered list processing more generally. Insofar as the latter two claims are supported, it would suggest a more nuanced account of the MNL hypothesis is required. To investigate these claims, 84 participants completed a NP and an alphabet-to-position task in which they marked the position of numbers/letters on a horizontal line. Of interest was whether: (1) similar judgment deviations from linearity occurred for number/letter stimuli; (2) left-to-right or right-to-left lines similarly, affected number/letter judgments; and (3) response times (RTs) differed as a function of number/letter stimuli and/or reverse/standard lines. While RTs were slower marking letter stimuli compared to number stimuli, they did not differ in the standard compared to the reverse number/letter lines. Furthermore, similar patterns of non-linear RTs were found marking stimuli on the number/letter lines, suggesting that similar strategic processes were at play. These findings suggest that a general mental representation may underlie ordered list processing and that a linear mental representation is not a unique feature of number *per se*. This is consistent with the hypothesis that number is supported by a representation that lends itself to processing ordered sequences in general.

**Keywords:** ordered sequences, non-numerical order, ordinal information, number representation, number line, numerical estimation

## INTRODUCTION

Francis Galton (1880) was an early advocate of the position that space and number are related. More recently, it has been proposed that the brain represents numerical magnitude information on something akin to a mental number line (MNL; Dehaene, 2011). The MNL model suggests numbers are represented in ascending order from left-to-right in a “continuous, quantity-based, analogical format” (Zorzi et al., 2002, p. 138). Inferences about the nature of the MNL have been mostly based on data from a number-to-position (NP) task (see Siegler and Opfer, 2003), as well as studies on

the spatial-numerical association of response codes (Dehaene et al., 1990, 1993), and spatial neglect (Zorzi et al., 2002).

In the NP task, participants mark the position of numerical values on a physical number line on which only the numerical end points are typically specified (e.g., ‘1’ and ‘100’). Findings show a close alignment (i.e., a linear algebraic function) between the marked positions of numerical values and their actual positions in older children and adults, while younger children typically overestimate small numbers and underestimate large numbers, and whose performance is best characterized by a logarithmic algebraic function. These findings are often used to suggest that the logarithmic/linear function that best fits NP task performance also characterizes the underlying form of the MNL. However, this interpretation should be treated with caution for at least two reasons. Firstly, it ignores the possibility that other processing factors affect, or may be responsible for, NP judgments. Secondly, it assumes that the NP/MNL relationship is unique to number. Clarifying these two possibilities may provide a more nuanced account of the relationship between number and space.

Evidence, albeit developmental evidence, suggests that performance on the NP task may not be a direct measure of numerical representations, and may depend on other processing factors (Barth and Paladino, 2011). NP estimation patterns instead may reflect an increased use of landmarks (the strategy of dividing lines into halves or quarters to create reference points to guide estimates) associated with a proportion judgment model, rather than a change in number representations *per se*. Barth and Paladino (2011) showed that children’s NP estimates were better fit by a proportion judgment model, rather than a “logarithmic-to-linear shift” account. This finding implies that performance on the NP task is not a direct measure of numerical representations because task strategies may influence NP task performance. While Barth and Paladino (2011) suggest NP task performance may be better described by a proportion power function rather than logarithmic/linear functions, they do not address whether NP task performance provides information about the representation of number *per se*.

It is clear that NP task manipulations may affect NP judgments. Cohen and Blanc-Goldhammer (2011; see also Cohen and Sarnecka, 2014), for example, used an “unbounded” NP task in which the number line represents a single unit distance from 0 to 1, and number estimations are made external to the bounds instead of within the bounds. On each trial, adult participants were presented with a target number (e.g., 12), and told to move the right boundary to estimate the target location. When participants completed a standard “bounded” NP task, estimation patterns closely followed a proportion judgment model, however, when they completed the unbounded NP task (which removes proportion judgment strategy opportunities), performance was best characterized by a linear function. This shows that task performance may reflect an increase in number line measurement skills, rather than a change in the underlying representation of number on the MNL (see also Huber et al., 2014). It also raises the issue of whether performance on a NP task reflects the way in which ordered lists, not just ordered number lists, are learned. Indeed, there is evidence that the pattern of mapping on the NP task may not be unique to number.

If number/letter line tasks, for example, produced the same mapping relationship to space, it would imply that linearity on the NP task may reflect a general mapping between spatial direction and list order. Several studies have investigated this possibility (Gevers et al., 2003, 2004; Berteletti et al., 2012; Hurst et al., 2014). For example, Hurst et al. (2014) gave children and adults an alphabet-to-position (AP) task, with the end points ‘A’ and ‘Z’. Children showed a logarithmic-to-linear shift with letters, and adults’ performance was characterized by a linear equation. While Hurst et al. (2014) were the first to show that numbers/letters spatial mapping abilities could be represented by linear/logarithmic functions, it is possible that performance on their task could be also represented by other algebraic functions (see comments on Barth and Paladino, 2011, above). Insofar as number/letter symbols display similar patterns on the spatial mapping task, it would argue against the claim that judgment patterns in the NP task are specific to number. Nevertheless, the issue of how best to compare similar/different number/letter judgment patterns, and *ipso facto* potential indices of different mental representations, remains an open question. In the present paper we address this issue by examining similarities/differences in non-linear deviations in NP and AP task performance.

In short, it is unclear whether NP judgment patterns are unique to number, or a general property of ordered lists that can be arranged spatially (e.g., the position of letters in an alphabet). To investigate this issue, we compared adults’ responses on NP and AP tasks, the aim of which was to determine whether the hypothetical mental representation of these two types of ordered lists differed. The research was designed to answer whether: (1) similar judgment deviations from linearity occurred for number/letter stimuli; (2) left-to-right or right-to-left lines similarly, affected number/letter judgments; and (3) response times (RTs) differed as a function of number/letter stimuli and/or reverse/standard lines. If patterns of spatial mapping and RT are similar for NP and AP tasks, it would suggest that NP task performance may reflect the way ordinal lists are processed, rather than anything specific about how numbers are represented. Such evidence would cast doubt on the value of drawing unique inferences about the underlying representation of number from spatial mapping tasks. Alternatively, if numbers/letters are found to have independent spatial mapping or RT patterns on NP and AP tasks, this would suggest that number symbols involve unique representations that can be studied using the NP task.

## MATERIALS AND METHODS

### Participants

Eighty-four ( $M = 19.05$  years,  $SD = 2.72$  years; 30 males, 54 female) undergraduate students from an Australian university participated in the research for course credit (our sample size was determined by how many participants we were able to recruit via our first year research participation program, and is a similar size to the sample used by Berteletti et al., 2012). All participants had normal or corrected-to-normal vision. Written and informed consent was obtained by asking participants to read a plain language statement and sign a consent form. All procedures

involved were approved by the University of Melbourne Human Ethics Advisory Group (HREC number 1441499).

## Apparatus

Stimuli were created on a Dell OptiPlex 9020 computer running Ubuntu with MATLAB software and Psychophysics Toolbox routines (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), and displayed on a 23 inch Dell E2314H LED monitor operating at a spatial resolution of 1,920 by 1,080 pixels at a refresh rate of 60 Hz.

## Stimuli

Participants completed a horizontal NP estimation task using integers between '1' and '26', and an alphabet-to-position (AP) task using letters between 'A' and 'Z' (with list positions that matched the numbers). In the NP task, the line endpoints were anchored with '1' on the left and '26' on the right, or '26' on the left and '1' on the right (i.e., a reversed NP). Participants positioned a target number (2 to 25) along the line to indicate a location in space that corresponded to the number's numerical position relative to the numerical endpoints. In the AP task, the lines were anchored with 'A' on the left and 'Z' on the right, or 'Z' on the left and 'A' on the right (i.e., a reversed AP). Participants positioned a target letter (B to Y) along the line to indicate a location in space that corresponded to the letter's alphabetical position relative to the letter endpoints.

Participants were tested on half the possible items in the number/letter lists to minimize the total number of trials. Numbers/letters were chosen by selecting every second number and the corresponding list position letter between 2 and 25 (e.g., '3' and 'C'). Participants were randomly assigned to even number/letter or odd number/letter target symbol conditions. The even condition set used the twelve target numbers: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, and the corresponding letters: B, D, F, H, J, L, N, P, R, T, V, and X. The odd condition participants were shown twelve target numbers: 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, and 25, and the corresponding target letters: C, E, G, I, K, M, O, Q, S, U, W, and Y. The target letter was displayed 8 cm above the middle of the horizontal line. The horizontal line was 15.5 cm long and 0.1 cm wide. The vertical marker used to indicate the target spatial position was 0.8 cm long and 0.1 cm wide. All stimuli were presented in black on a white background. Number/letter stimuli were presented in Arial font about 1 cm high.

## Procedure

Testing was conducted in a quiet testing space with participants seated about 60 cm in front of the computer monitor. Participants were presented with a series of horizontal lines on the screen, anchored by a number/letter on the left and right ends of the line. On each trial a number/letter appeared in the middle of the screen, above the line. Participants were instructed to move the computer mouse (i.e., vertical marker) left or right to the place where they thought the symbol should be on each line. Participants were asked to move the mouse marker and click when they thought they had positioned the symbol correctly.

Trials were randomized within a block of 48 trials (numbers/letters, left-to-right or right-to-left, 12 different targets), and participants completed nine blocks (i.e., 432 trials in total). Participants had as long as they wished on each trial to make their response. Trials were separated by 500 ms to reduce the potential for outlier bias in calculating RTs, participants with fewer than seven repeats for any condition were excluded from the analyses. Eighteen participants were excluded using this criterion, resulting in a final sample size of 66 participants, 79.57% of whom completed the testing.

## Analytic Approach

The overall aim of the analyses was to investigate similarities/differences in possible nonlinearities in the NP and AP data. As a first step though, we report the linear/logarithmic equation fits to our data. Specifically, we report the mean logarithmic and linear fit residuals for the NP and AP tasks to provide an overview of the data. To determine whether performance differed for the NP and AP tasks, we compared the spatial mapping RTs for number/letter symbols for the left-to-right version of the task, as well as right-to-left version of the task. The aim was to identify possible nonlinearities in the different tasks.

To more closely examine possible nonlinearities in these data, we plotted the average residuals from a linear fit for numbers/letters for both directions of the task. And to determine whether nonlinearities in the residuals depended on symbol type or task direction, we compared performance across numbers/letters and task direction. We also compared RT data for numbers/letters in both task directions to examine strategies used for NP/AP tasks.

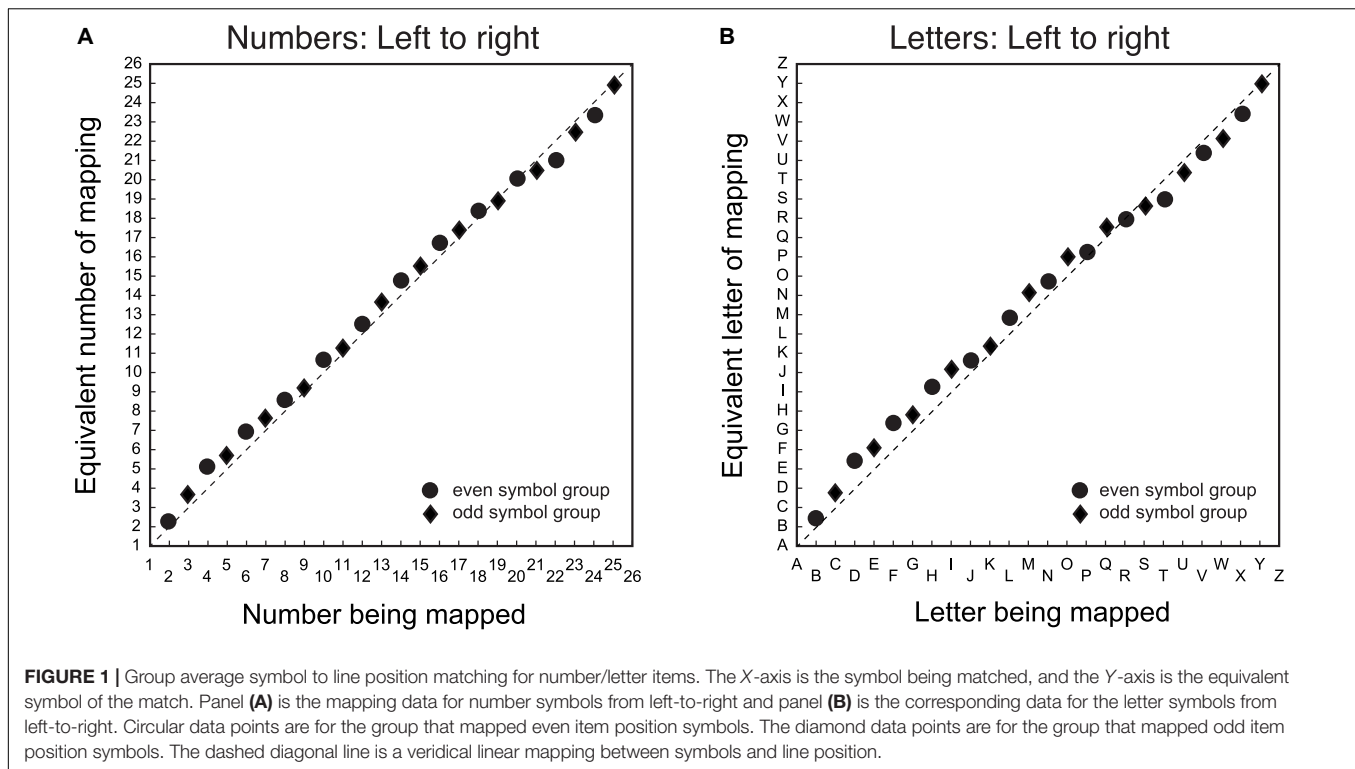
To achieve the last three steps, we used a nested bootstrapping method to generate 95% confidence intervals. A distribution of group mean data was generated by calculating the mean of 10,000 samples of participants (with replacement) using resampled raw mapping data (with replacement) for each participant. To get a better idea of the underlying patterns in these data, we also fit polynomial curves to the data using the `polyfit` function in MATLAB.

## RESULTS

Overall, the data are consistent with a linear mapping function. For each participant in each condition, we calculated the mean of the residuals from both a linear and a logarithmic mapping function as a proportion of the line length. In every case, the mean linear residual was lower than the mean logarithmic residual. The group average mean linear residual was 0.005, whereas the group average mean logarithmic residual was 0.064. This shows that a logarithmic model was not a good fit for any participant in any of the conditions, and that a linear model was a better general characterization of task performance.

## Mapping Between Symbols and Space

The first step was to determine whether there was a difference between number/letter symbols in the spatial mapping task.



**Figure 1** shows the group mean average symbol to line position matching for number (**Figure 1A**) and letter (**Figure 1B**) items. The solid symbols are the group averages of individual mapping data, which is the mean of 6 to 9 repetitions of each item. The data deviated minimally from the diagonal (the average residuals were typically less than 5% of the spatial distance between items).

It is evident that there is a general linear mapping between symbols and space that does not depend on symbol type. While the pattern of the mapping between symbol and space is similar for both numbers/letters, the systematic deviations from the diagonal suggest that these mapping functions both have non-linear components. Specifically, there was a tendency to mark the line further to the center of the veridical linear location of the symbol for items proximal to, but not immediately adjacent to, the end points. There was also a tendency for items in the middle of the list to be positioned toward the right of the veridical location.

Non-linear mappings of number to space have been reported previously (e.g., Siegler and Opfer, 2003). However, these nonlinearities are typically characterized as logarithmic or quasi-logarithmic patterns, where symbols are mapped toward the right of the linear prediction for all items. The present results differ from previously reported logarithmic patterns in two ways. Firstly, the nonlinearities are small. Secondly, the patterns of non-linear residuals, both above and below a veridical linear function, are not consistent with a mapping model based on a power or quasi-logarithmic function. It is possible the systematic nonlinearities in the spatial mapping function for numbers/letters may reflect a bias to either under- or over-estimate the position of number/letter items due to the properties

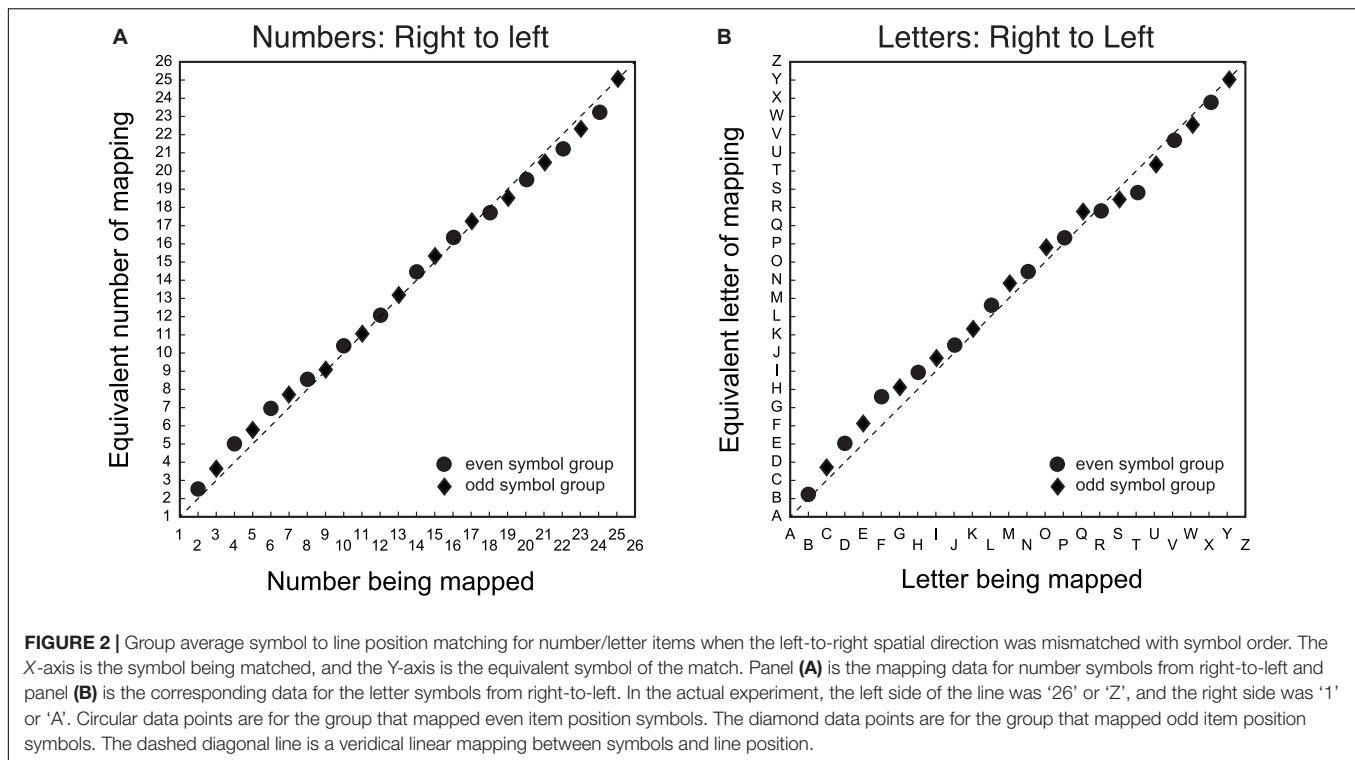
of the task. Alternatively, the pattern of nonlinearities may reflect list order effects common to both numbers/letters when they are mapped to space.

## Mapping Between Symbols and Space With Reversed Mapping Direction

The second step was to determine whether there was a difference between number/letter symbols in the spatial mapping task when the mapping direction was switched from left-to-right to right-to-left. **Figure 2** displays the group mean average symbol to line position matching in the mapping task for number (**Figure 2A**) and letter (**Figure 2B**) items, where the spatial direction of the line was reversed by anchoring the left side of the line with '26' or 'Z', and the right side with '1' or 'A', respectively. The data are displayed in the item order rather than the spatial direction, because consistent patterns in the data depend on item position. As with **Figure 1**, the mapping between number symbols and line position deviates minimally from the diagonal, consistent with a linear mapping from number symbols to space.

Also similar to **Figures 1, 2** shows systematic deviations above and below the diagonal that are qualitatively similar for both number/letter symbols. There is a tendency to map symbols onto the line further to the center of the veridical linear location of the symbol. Since the direction of the line was dissociated from item order, nonlinearities in the same direction for **Figures 1, 2** reflect nonlinearities that are related to item order, whereas nonlinearities in the opposite direction are nonlinearities that depend on spatial direction.





## Group Average Residuals of Symbol to Line Mapping

The third step was to examine the group average residuals from the veridical location for the number/letter symbols in the spatial mapping task. **Figure 3** shows that the residuals from the veridical linear are qualitatively similar for symbol type and mapping direction. **Figures 3A,B** show the data for numbers. **Figures 3C,D** show the data for letters. **Figures 3A,C** show the data for left-to-right item direction. **Figures 3B,D** show the data for right-to-left item direction. To determine if the residuals systematically deviate from a linear mapping function, a sixth order polynomial was fit to the 95% confidence intervals (CIs) using polyfit in MATLAB.

These data show that there is a tendency to map symbols that are a fourth or three-fourths of the way through the item list length toward the center item. The pattern appears to be systematic, given that all data display this trend regardless of symbol or mapping direction. Furthermore, given that the polynomial fit to the 95% CIs do not include 0, the non-linear pattern near the end points is unlikely to have occurred by chance.

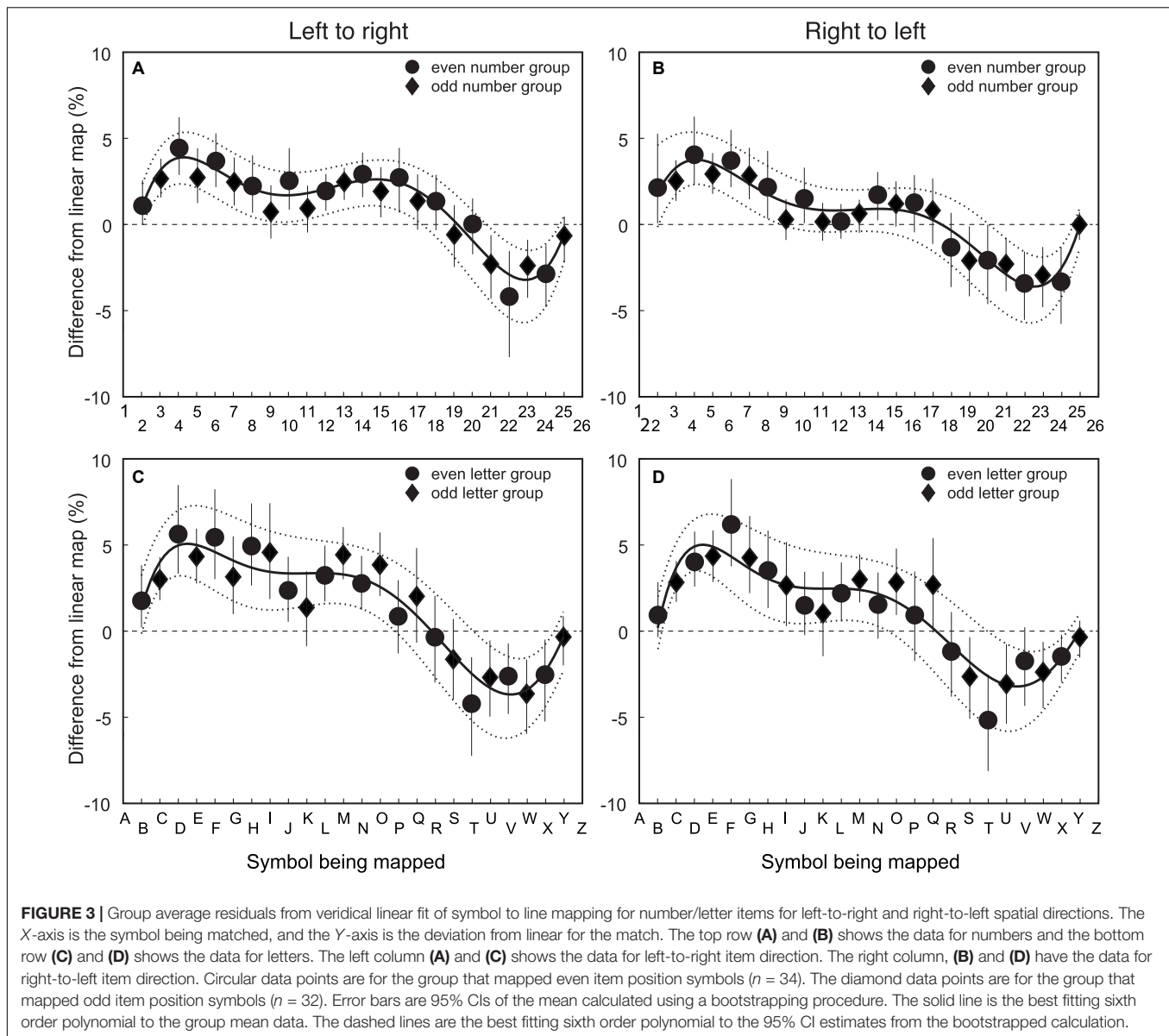
The data in **Figure 3** show a small systematic tendency for items toward the middle of the list to be positioned away from the start of the list. This is evident for both numbers/letters. The bias is larger for left-to-right ordered items, suggesting that this effect is strongest when the item order and spatial mapping direction are matched. The polynomial fit to the 95% CIs for the left-to-right mapping data provides support for the proposition that the bias is unlikely to have occurred by chance. The polynomial fit to the 95% CIs for the right-to-left letter data does not include

0, meaning the residuals represents a systematic bias. While the polynomial fit to 95% CIs for the right-to-left number data does not provide evidence of systematic bias, there is a trend in the same direction as the other conditions. The pattern of results suggests there are few differences in the nonlinearities across symbol type or mapping direction.

## Comparison of Mapping for Symbol Type and Mapping Direction

The fourth step was to compare spatial mapping across number/letter symbols and mapping direction. **Figure 4** displays the *t*-score for each item position. **Figure 4A** shows mapping direction for number symbols. **Figure 4B** shows mapping direction for letter symbols. **Figure 4C** shows symbol type for left-to-right mapping. **Figure 4D** shows symbol type for right-to-left mapping. To address how much the residual patterns depend on symbol type and the congruency between symbol item order and spatial direction, we performed a series of permutation *t*-tests for each item position, comparing mapping direction and symbol type. Comparisons of item direction were performed separately for each possible condition. In the permutation *t*-test, the 95% CI of the *t*-statistic for each data set was calculated using a bootstrapping procedure. The distribution of *t*-statistics was obtained by re-computing the *t*-value 10,000 times with random assignment of the data to each group. If the *t*-value of the data is outside the 95% CI for the bootstrap *t*-distribution, then we can infer that the data are different for the condition.

The trend in *t*-scores suggests that number mapping differs for left-to-right versus right-to-left direction for symbols in the middle of the list. There is a similar trend for letters, but the



$t$ -scores for letters are not as large as the number  $t$ -scores, or as consistently outside the 95% CIs. Although the  $t$ -statistics show that there are systematic differences in number mapping for left-to-right versus right-to-left, the magnitude of the bias is small. The large  $t$ -values for middle items are partly due to less variance (or greater consistency) of position data across individuals for middle number items. As such, there may be more precision in the estimates for numbers (and less so for letters) in the middle of the list.

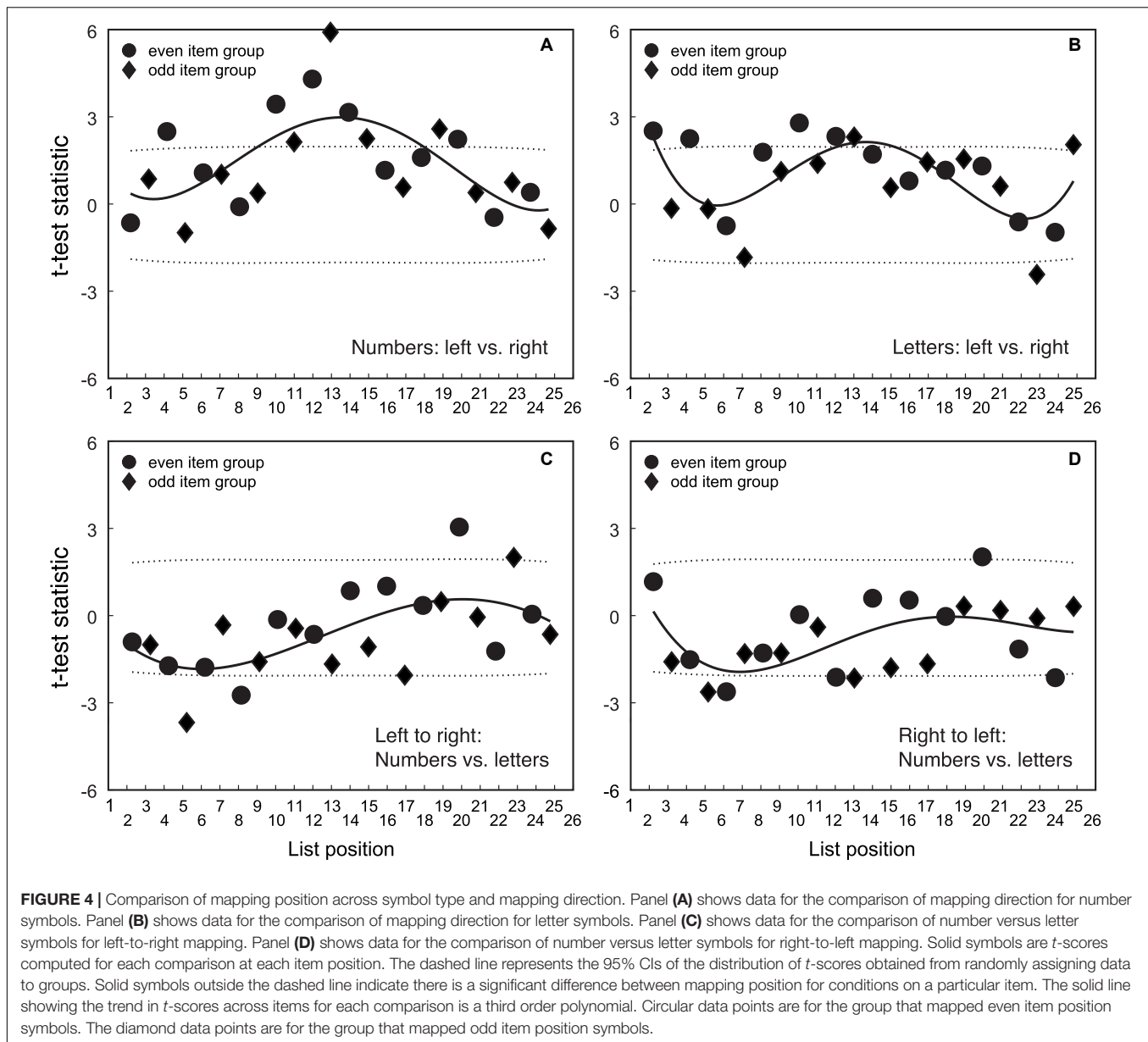
In sum, the pattern of  $t$ -statistics suggests that left-to-right and right-to-left mapping is similar for numbers/letters. There is a small trend for larger differences between numbers/letters in position mapping for symbols at the top of the item list. However, there is little in the position mapping data that reveals differences in the spatial mapping of numbers/letters. Nevertheless, there are differences in the response times (RTs)

taken to perform symbol to space mapping which may be clarified with further analysis.

## Response Times for Symbol Type and Mapping Direction

The fifth step was to compare RTs across number/letter symbols and mapping direction. Figure 5 shows group average RTs for matching symbols to line position. Figure 5A shows the RT to map numbers that are ordered left-to-right. Figure 5B shows the RT to map numbers that are ordered right-to-left. Figure 5C shows the RT to map letters that are ordered left-to-right. Figure 5D shows the RT to map letters that are ordered right-to-left.

The RT to map a symbol onto a line has a distinct pattern as a function of item position that does not depend on symbol type



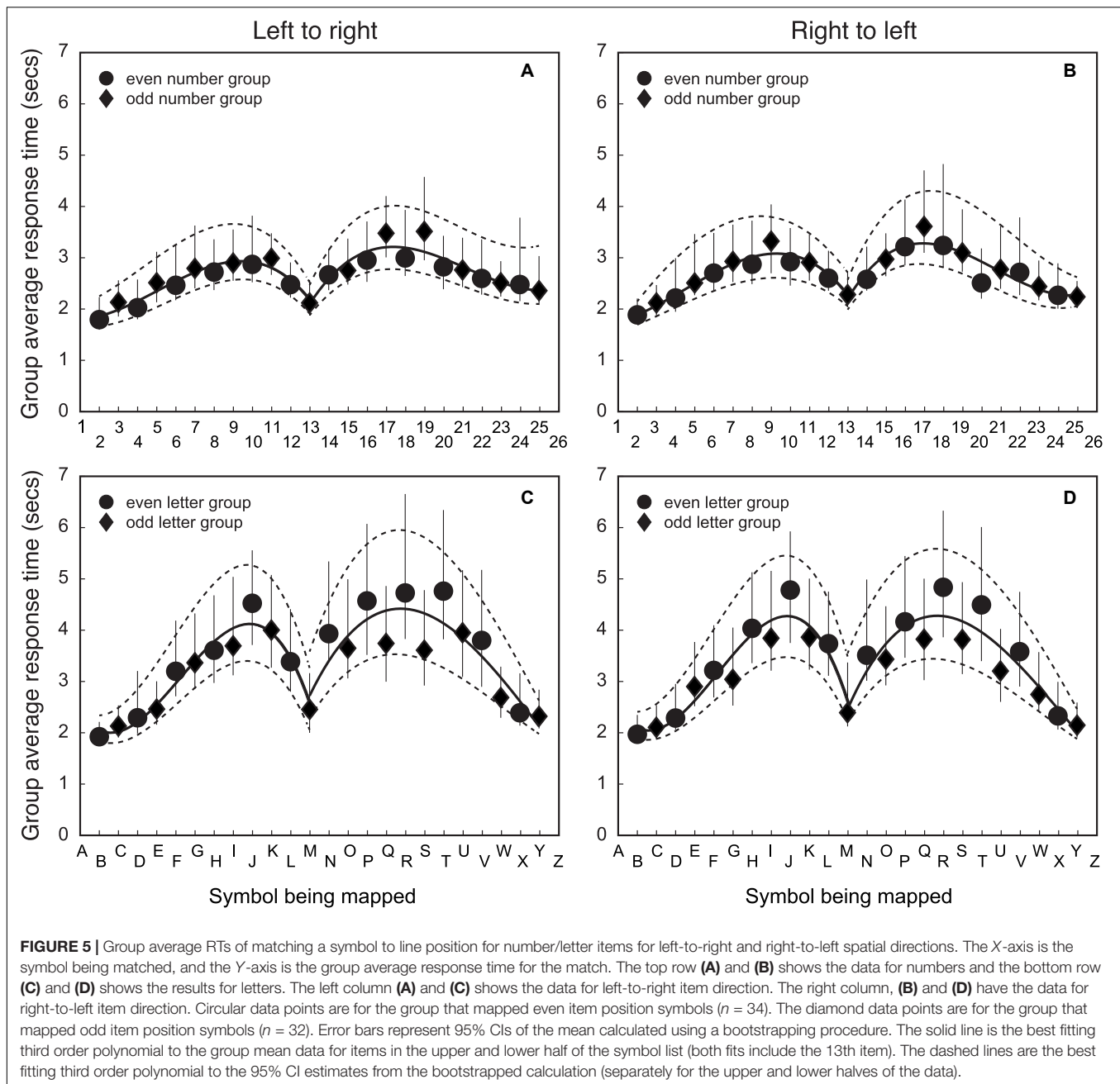
or mapping direction. Mapping is fastest for items at either end of the list, becoming slower for items further from the list end, with RTs being slowest for items about a third from the end. RTs then decrease for symbols in the middle of the list, with the 13th element of the list ('13' or 'M') being mapped almost as quickly as items at the ends of the list. Given that '13' is half the list length of '26' if counting from the top of the list, the RT data suggest that symbol list order is important for understanding the process of how symbols are mapped onto the line position.

The pattern of RTs does not seem to depend strongly on mapping order. The magnitude and pattern of change in RTs with item position in panel 5A is very similar to panel 5B. Panel 5C is also similar to panel 5D. The importance of the 13th item for numbers/letters suggests that item order is processed in a similar way regardless of the symbol type. However, the variability in

letter RTs is much larger than that for number RTs, so a small effect would be less evident in these data.

Other patterns of RTs may provide information about how participants map number/letter symbols to spatial locations on a line. The RTs for 'B', 'C' and 'D' are similar to '2', '3' and '4'. Similarly, the speed for mapping 'X' and 'Y' is similar to '24' and '25'. The RT variance is also smaller for these letters compared to those far from the end of the list. This is good evidence of a processing speed advantage when item positions are close to the beginning or end of the list. Such positions are also easier to calculate for letters. The letter items toward the middle of the list take longer to calculate, and this may be reflected in the longer RTs for letters toward the middle of the list compared to numbers.

Overall, the RT patterns indicate that the strategies used to map numbers to spatial position are similar to that used to map



**FIGURE 5 |** Group average RTs of matching a symbol to line position for number/letter items for left-to-right and right-to-left spatial directions. The X-axis is the symbol being matched, and the Y-axis is the group average response time for the match. The top row (A) and (B) shows the data for numbers and the bottom row (C) and (D) shows the results for letters. The left column (A) and (C) shows the data for left-to-right item direction. The right column, (B) and (D) have the data for right-to-left item direction. Circular data points are for the group that mapped even item position symbols ( $n = 34$ ). The diamond data points are for the group that mapped odd item position symbols ( $n = 32$ ). Error bars represent 95% CIs of the mean calculated using a bootstrapping procedure. The solid line is the best fitting third order polynomial to the group mean data for items in the upper and lower half of the symbol list (both fits include the 13th item). The dashed lines are the best fitting third order polynomial to the 95% CI estimates from the bootstrapped calculation (separately for the upper and lower halves of the data).

letters to spatial position. The mapping process requires finding the item position, converting that to a proportion of list length, and mapping that into visual space. Such a process does not require there to be a spatial arrangement of numbers along a number line. However, we now have evidence that numerical ability is necessary to complete a spatial mapping task with numbers/letters. This numerical requirement supports the idea that changes in the spatial mapping of numbers onto a line in the NP task may reflect increased familiarity with ordered lists, the speed an item position can be determined, and the ability to calculate a proportion from the item position and total list lengths.

## DISCUSSION

We compared the pattern of responses for NP and AP tasks. Our aim was to examine whether mapping of number to space, as indexed by the pattern of responses in a NP task, is unique to number, or reflects the way in which ordered lists are spatially represented (i.e., the AP task). Of interest was whether: (1) similar judgment deviations from linearity occurred for number/letter stimuli; (2) left-to-right or right-to-left lines similarly, affected number/letter judgments; and (3) RTs differed as a function of number/letter stimuli and/or reverse/standard lines. Three findings are of note. Firstly, similar deviations



from linearity were found for numbers/letters, suggesting that participants used similar methods for judging the position of symbols in the two tasks. Secondly, end point symbols did not affect performance. Thirdly, participants took longer to make AP compared to NP task judgments. These findings support the view that NP judgments may not be uniquely informative about the MNL because the pattern of performance was similar for number/letter symbols.

Our findings are consistent with those of other researchers who have argued that changes in NP judgment patterns reflect an increased use of proportion judgment strategies (Barth and Paladino, 2011), or increased list processing automaticity (Hurst et al., 2014). We also found systematic nonlinearities in both number/letter mapping patterns that may reflect a small bias for marking lines away from end points. These small linear deviations are not explained by combinations of linear and/or logarithmic mapping functions, which would produce only positive residuals. Our data do not support the view that mapping symbols onto a horizontal line depends uniquely on the numerical properties of the symbols associated with their location on a hypothetical MNL. Rather, our findings are consistent with a model in which the symbol position within an ordered list is used to map a symbol to a location on a horizontal line.

We found little evidence that mismatches in number/letter symbol position order, or task direction, alter the shape of the mapping function between symbols and space. Specifically, the patterns of non-linear residuals were similar regardless of symbol type and regardless of whether the spatial direction of the horizontal line was matched with numerical magnitude direction. We found a small trend for middle items to be estimated toward the right side of space (an effect that was more evident for numbers than letters), but the *t*-statistics for combinations of symbol type and direction congruence show that the patterns were similar across conditions. These findings are consistent with the claim that participants use a single method to map the position of number/letter symbols to the location along the horizontal line.

The RT data showed that responses were faster for items near the end of the list and close to the middle of the item sequence. This indicates participants were most efficient judging the position of symbols that reflected a fraction of a half. While letter judgments were slower than number judgments, when the item position of the letter is known, RTs for the comparable symbol sets was similar. For example, the relatively fast RTs of '13' and 'M' suggest that it is more efficient to map symbols that map onto simple fractions. This is consistent with the view that mapping the line position of symbols requires the use of list length to calculate the position of an item as a proportion. This calculation does not depend on the line end points. In other words, our findings are consistent with a model that suggests that the ability to judge proportions is critical to performing the NP task.

We suggest that the linearity of judgments on NP tasks may reflect how ordered lists are learned (items in the beginning of number/letter sequences are learned before elements later in the list), rather than unique information about the MNL. This is because there is no numerical requirement for letters, which are

items that are ordered without inherent magnitude (e.g.,  $1 + 2$  is meaningful, whereas  $A + B$  is not), to be spaced linearly with a fixed interval between each item. Our findings support those of Hurst et al. (2014) and Berteletti et al. (2012), who show that older children and adults mark letters linearly on an AP task. They cast doubt on the value of making inferences about NP task performance as a measure of the underlying number representation. Specifically, the mapping process in NP judgments may not involve decoding numerical information from a MNL and suggests that NP judgments may reflect list order learning of numbers in formal education.

It is important to note that NP judgments may require "number" precision, irrespective of the elements being mapped. A participant must know where an item is in a list, be able to calculate the relative position in the list, and be able to match the relative proportionate location of items on the line. In other words, participants must be able to calculate relative proportions of items on the line. For numbers/letters, participants were faster and more accurate for items at the beginning and end of lines, and for items in the middle of the line, most likely because it was easier to calculate these proportions. If a participant diagnosed with dyscalculia (Butterworth et al., 2011) were to complete this task, for example, we would expect equally poor performance for the number/letter elements, since this participant might be incapable of calculating the relative proportions of the line, even if they knew the order of the letters. The fact that letter responses were slower overall suggests participants may be mentally mapping letters onto numbers to complete the task.

A possible limitation of our study is that we analyzed group data without regard to the possibility of individual differences. However, given that the standard errors were low, we think there were probably few systematic differences in mapping patterns. Nevertheless, future work should investigate whether individual differences are related to the ability to estimate item position and make proportion calculations. Furthermore, evidence that familiar letters near the end of the list were processed as quickly as numbers suggest that list familiarity may be important in shaping the methods participants use in line judgment tasks. Future work with artificial non-numerical symbols could reveal learning processes that are relevant to numerical learning more generally.

Overall, our findings support the claim that the NP task should not be taken as a unique measure of number along a hypothetical MNL. Our results are consistent with the idea that ordinality plays an important role in explaining the mapping function on the NP task, rather than the magnitude of number symbols being matched. It seems possible that instead of a number specific representation, number may be supported by a representation that is able to process ordered lists in general.

## AUTHOR CONTRIBUTIONS

CP contributed to investigation, methodology, data collection, and editing the original draft. RR contributed to draft editing and supervision. JF prepared the methodology, coding, data analysis, draft editing, and supervision.

## REFERENCES

- Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x
- Berteletti, I., Lucangeli, D., and Zorzi, M. (2012). Representation of numerical and non-numerical order in children. *Cognition* 124, 304–313. doi: 10.1016/j.cognition.2012.05.015
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897x00357
- Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536
- Cohen, D. J., and Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychon. Bull. Rev.* 18, 331–338. doi: 10.3758/s13423-011-0059-z
- Cohen, D. J., and Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Dev. Psychol.* 50, 1640–1652. doi: 10.1037/a0035901
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122:371. doi: 10.1037/0096-3445.122.3.371
- Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 16:626. doi: 10.1037/0096-1523.16.3.626
- Galton, F. (1880). Visualised numerals. *Nature* 21, 252–256. doi: 10.1038/021252a0
- 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., Reynvoet, B., and Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: evidence from days of the week. *Cortex* 40, 171–172. doi: 10.1016/S0010-9452(08)70938-9
- Huber, S., Moeller, K., and Nuerk, H. C. (2014). Dissociating number line estimations from underlying numerical representations. *Q. J. Exp. Psychol.* 67, 991–1003. doi: 10.1080/17470218.2013.838974
- Hurst, M., Leigh Monahan, K., Heller, E., and Cordes, S. (2014). 123s and ABCs: developmental shifts in logarithmic-to-linear responding reflect fluency with sequence values. *Dev. Sci.* 17, 892–904. doi: 10.1111/desc.12165
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in psychtoolbox-3. *Perception* 36, 1–16.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897x00366
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Zorzi, M., Priftis, K., and Umiltà, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

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

Copyright © 2019 Podwysocki, Reeve and Forte. 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.



# Individual Differences in Implicit and Explicit Spatial Processing of Fractions

Elizabeth Y. Toomarian\*, Rui Meng and Edward M. Hubbard

Department of Educational Psychology, University of Wisconsin–Madison, Madison, WI, United States

## OPEN ACCESS

### Edited by:

Firat Soylu,  
The University of Alabama,  
United States

### Reviewed by:

Robert Reeve,  
The University of Melbourne, Australia  
Mauro Murgia,  
University of Trieste, Italy

### \*Correspondence:

Elizabeth Y. Toomarian  
etoomarian@stanford.edu

### Specialty section:

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

**Received:** 15 September 2018

**Accepted:** 04 March 2019

**Published:** 05 April 2019

### Citation:

Toomarian EY, Meng R and  
Hubbard EM (2019) Individual  
Differences in Implicit and Explicit  
Spatial Processing of Fractions.  
Front. Psychol. 10:596.  
doi: 10.3389/fpsyg.2019.00596

Recent studies have explored the foundations of mathematical skills by linking basic numerical processes to formal tests of mathematics achievement. Of particular interest is the relationship between spatial-numerical associations—specifically, the Spatial Numerical Association of Response Codes (SNARC) effect—and various measures of math ability. Thus far, studies investigating this relationship have yielded inconsistent results. Here, we investigate how individual implicit and explicit spatial representations of fractions relate to fraction knowledge and other formal measures of math achievement. Adult participants ( $n = 105$ ) compared the magnitude of single digit, irreducible fractions to  $1/2$ , a task that has previously produced a reliable SNARC effect. We observed a significant group-level SNARC effect based on overall fraction magnitude, with notable individual variability. While individual SNARC effects were correlated with performance on a fraction number-line estimation (NLE) task, only NLE significantly predicted scores on a fractions test and basic standardized math test, even after controlling for IQ, mean accuracy, and mean reaction time. This suggests that—for fractions—working with an explicit number line is a stronger predictor of math ability than implicit number line processing. Neither individual SNARC effects nor NLE performance were significant predictors of algebra scores; thus, the mental number line may not be as readily recruited during higher-order mathematical concepts, but rather may be a foundation for thinking about simpler problems involving rational magnitudes. These results not only characterize the variability in adults' mental representations of fractions, but also detail the relative contributions of implicit (SNARC) and explicit (NLE) spatial representations of fractions to formal math skills.

**Keywords:** spatial-numerical associations, SNARC, number line estimation, fractions, individual differences

## INTRODUCTION

Recent efforts to understand predictors of mathematical achievement have begun to focus on the contribution of spatial skills in addition to numerical abilities. This initiative has widespread educational implications, as spatial ability in early teenage years predicts the eventual likelihood of pursuing advanced study in STEM (Science, Technology, Engineering and Mathematics) topics and careers in a STEM field (Shea et al., 2001; Wai et al., 2009). The combined development

of spatial and numeracy skills are unique predictors of later mathematical success and other academic outcomes, with strong cross-domain links evident from early childhood (for a review, see Mix and Cheng, 2012). For instance, spatial skills at age 5 have been shown to predict standardized math scores at age 7 (Gunderson et al., 2012; Gilligan et al., 2017), and a number of spatial skills (e.g., mental rotation, visuospatial working memory) predict math performance throughout childhood. One possible account for these relationships is the close behavioral, cognitive, and neural link between numbers and space (e.g., Hubbard et al., 2005; Toomarian and Hubbard, 2018a).

These findings highlight just a few of the many factors that contribute to early mathematical understanding. Multiple numerical abilities likely serve as precursors to greater mathematical ability, though some may contribute more or less than others, with many competencies being closely related. For instance, in one specific study, preschool children's approximate number sense and cardinality knowledge of number words both predicted later math achievement, and cardinality was found to mediate the relationship between approximate number and math achievement (Chu et al., 2015). Further investigation of these factors is certainly needed, particularly as they relate to classes of numbers such as fractions, which are believed to be a critical part of a strong foundation for numerical understanding and uniquely predictive of later algebra-readiness (Booth and Newton, 2012).

In the current study, we specifically investigated the relationship between measures that link spatial and numerical processing of fractions by using several measures of implicit and explicit spatial-numerical associations (SNAs). We then aimed to determine the unique contribution of these factors to multiple measures of formal math achievement, such as tests of fractions arithmetic and algebra.

## Spatial-Numerical Associations and the Link to Mathematics

Spatial and numerical cognition have been studied in conjunction since at least the 19th century (Galton, 1880), with mounting evidence that both evolutionary and cultural factors contribute to the widely-evidenced link between the two (for a review, see Toomarian and Hubbard, 2018a). The link between numbers and space is supported from a number of theoretical perspectives. The mental number line (MNL) theory suggests that people have an internal representation of a number line, along which numerical magnitudes extend horizontally in the direction congruent with their primary written language (e.g., left-to-right for English readers) (Dehaene et al., 1993). This internal conceptualization links numbers and space along a linear continuum. There is also theoretical support from a developmental perspective; one of the central claims of the integrated theory of numerical development (Siegler et al., 2011) is that solid mathematical understanding requires knowing that all numbers have magnitudes that can be spatially oriented and placed on number lines. Despite the theoretical basis for a link between spatial skills and numerical cognition, it is unclear whether SNAs directly influence complex cognitive functions such as mathematical thinking.

In order to measure the implicit link between numbers and space, researchers typically employ one of several behavioral tasks, the most common being a parity or numerical judgment task with spatially-coded responses. In the magnitude judgment task, participants indicate whether a number is larger or smaller than a standard reference number by using either a left- or right-side response key, while in the parity task participants indicate whether the given number is even or odd. Dehaene et al. (1993) were the first to demonstrate that people were consistently faster to respond to relatively smaller magnitudes on the left and larger magnitudes on the right during parity judgment, a phenomenon termed the Spatial Numerical Association of Response Codes—or SNARC—effect. This response pattern is often taken as evidence of a MNL (Dehaene et al., 1993; Fias et al., 1996; Hubbard et al., 2009; but see Nuerk et al., 2015; Proctor and Xiong, 2015; Abrahamse et al., 2016 for recent discussion of alternative explanations). This effect has been demonstrated across many stimulus types (e.g., Nuerk et al., 2005; Ren et al., 2011; Prpic et al., 2018). Furthermore, the SNARC effect is generally viewed as an implicit, quantitative measure of a person's internal conception of spatially-oriented number and may prove to be useful in illuminating the building blocks of complex mathematical thinking. The distance effect, or the finding that numbers “closer” in numerical magnitude are more difficult to discriminate than those that are “farther” (Moyer and Landauer, 1967; Restle, 1970), is also often taken as evidence of a MNL, though it should be noted that this effect is not sensitive to spatial organization or direction.

The relationship between individual SNARC effects and formal mathematical abilities has become an emerging topic of interest, yet the nature of this relationship is still not well defined. Recent studies of the SNARC have highlighted notable variability in the strength and direction of people's SNARC effects. Despite group-level effects that indicate a classic SNARC effect, about 20–40% of individuals either have no SNARC effect or one that would suggest a right-to-left SNA (Wood et al., 2006; Cipora and Wood, 2017, Supplementary Material). Unfortunately, attempts to link this variability in SNAs to mathematical proficiency have yielded mostly paradoxical findings, with greater math skill related to weaker or null SNARC effects for whole numbers in adults (Cipora and Nuerk, 2013; Hoffmann et al., 2014) and children (Schneider et al., 2009; Gibson and Maurer, 2016).

However, there has been some evidence that spatial ability may account for these differences. Viarouge et al. (2014) demonstrated that individual differences in the whole number SNARC were explained by measures of spatial cognition and distance effects. Furthermore, a group of professional engineers exhibited significant SNARC effects, while expert mathematicians did not (Cipora et al., 2016; see also Hoffmann et al., 2014). This is further supported by a study of spatial representations of angle magnitude, with engineering students showing SNARC-like effects for angles whereas psychology students did not (Fumarola et al., 2016). This suggests that other factors, such as visuospatial/mental imagery skills or perhaps more domain-general skills rather than domain-specific ones, may be closely linked to the SNARC and



act as a mediating factor between MNL representations and math outcomes.

## Number Line Estimation and the Link to Mathematics

While the SNARC effect reveals an implicit link between numerical magnitudes and space, experimental paradigms using physical number lines attempt to more *explicitly* probe participants' underlying spatial conceptions of number. Perhaps the most common such paradigm is the Number Line Estimation (NLE) task, in which participants place a given number on a physical, horizontally-oriented line that typically includes labeled endpoints (e.g., Siegler and Opfer, 2003). Performance on the task is classically measured in terms of acuity and/or the linear fit of participant responses. This paradigm is widely used in the numerical cognition literature, as it provides a concrete link between physical and mental spatial representations of numerical magnitudes.

Several studies have now demonstrated a link between number line estimation ability and math achievement (Siegler and Opfer, 2003; Booth and Siegler, 2006; Muldoon et al., 2013; Friso-van den Bos et al., 2015; Simms et al., 2016), with greater acuity on NLE tasks associated with higher math ability. These findings have been validated by a recent developmental meta-analysis of such studies (Schneider et al., 2018), which found a strong correlation between number line estimation ability and measures of mathematical competence, including counting, arithmetic, school grades, and standardized test scores. The link between number line estimation and stronger internal magnitude representations has been extended to training studies using linear gameplay elements. Studies of board games that rely heavily on gameplay components reminiscent of number lines, such as *Chutes and Ladders*, have demonstrated a positive effect on a range of mathematically-relevant outcomes (Ramani and Siegler, 2008; Whyte and Bull, 2008; Siegler and Ramani, 2009), including numerical magnitude comparison, counting ability, and more formal number line estimation tasks.

Some scholars contend that the relationship between NLE performance and math proficiency can be attributed to other, related cognitive factors, many of which are spatial in nature. For instance, Simms et al. (2016) found that visuospatial abilities mediated the relationship between linearity of NLE responses and math achievement in children aged 8–10 years. Interestingly, Gunderson et al. (2012) found that number line performance mediated the relationship between spatial skills and early calculation abilities. Taken together, these studies point to the intertwined development of spatial ability and numerical estimation abilities underlying later math achievement.

## The Importance of Fractions

Notably, the entirety of this new research has focused solely on SNAs (and specifically the SNARC effects) for whole numbers. This is surprising, as recent behavioral studies have repeatedly demonstrated links between basic numerical abilities and individual differences in fraction knowledge. In middle school, fraction magnitude knowledge and whole number division have been shown to predict individual differences in both fraction

arithmetic and standardized math test scores (Siegler and Pyke, 2013). Furthermore, high-achieving students are more likely to rely on overall (holistic) fraction magnitude when doing fraction tasks, while low achievers are more likely to focus on the components, supporting the hypothesis that stronger holistic mental representations of fraction magnitudes leads to higher levels of overall math achievement (for similar evidence related to math learning disabilities, see Mazzocco et al., 2013). DeWolf et al. (2015) demonstrated that measures of *relational* fraction knowledge and placing decimals onto number lines were the best predictors of algebra performance. The predictiveness of relational fraction concepts may be supported by an underlying ratio-processing system (RPS), which is sensitive to non-symbolic ratios such as line length comparisons (Lewis et al., 2015). Acuity of the RPS is also related to formal math achievement, including performance on symbolic fraction tasks and algebra achievement scores (Matthews et al., 2016), bolstering the claim that holistic fraction magnitude processing is key for later math learning.

As evidence emerges that fractions provide a foundation for later achievement in mathematics, researchers have also begun to investigate the developmental predictors of elementary school children's fraction knowledge. A longitudinal study by Ye et al. (2016) demonstrated the importance of number line estimation, division and multiplication with whole numbers, as well as non-symbolic proportional reasoning, on later fraction knowledge. Additionally, Schneider et al. (2018) found that the relationship between NLE and math achievement became stronger with age, a pattern that could be attributed to fraction knowledge. Jordan et al. (2013) found that performance on a number line estimation task was the *largest* independent contributor to both conceptual and procedural fraction knowledge, highlighting the importance of SNAs for fraction understanding. As a number line estimation task is essentially an explicit measure of internal representations of the number line, this finding indicates that an implicit measure of SNAs (e.g., the fraction SNARC) might be similarly sensitive.

In line with this prediction and previous work on the SNARC effect for whole numbers, fractions have indeed elicited a group-level classic SNARC effect (Toomarian and Hubbard, 2018b). Inasmuch as whole number SNAs may be related to spatial or math-related outcomes, inter-individual variability in the fractions SNARC may be an important signature of differences in holistic fraction processing and mathematics ability more broadly. However, the link between the fraction SNARC and individual differences in math achievement has not yet been explored. Furthermore, no studies have investigated the possibility that a more explicit number line estimation task may mediate the relationship between the implicit fractions SNARC effect and spatial/mathematical measures. While Schneider et al. (2009) found that a parity based SNARC effect for whole numbers did not predict conceptual knowledge of decimal fractions and that a decimal NLE task did, it is unclear whether these findings would hold if fractions were used to elicit a SNARC instead. An independent effect of the fractions SNARC on mathematical outcome measures would further support the critical role of spatial processing in fraction processing and proportional reasoning (Möhrling et al., 2015).

## The Present Study

This study aimed to investigate the link between implicit spatial representations of fractions in adults and explicit measures of numerical/mathematical knowledge by focusing on three central questions: (1) which factors predict individual differences in spatial representations of fractions? (2) to what extent is the SNARC effect distinct from other indices of numerical processing (e.g., the distance effect and number line estimation) and (3) do spatial representations of fractions, as measured by the fractions SNARC and NLE task, uniquely account for differences in math achievement in university undergraduates?

With respect to the first two research questions, our predictions were largely influenced by theoretical considerations. If people consistently rely on the MNL when comparing numerical magnitudes, that would imply (1) that SNARC effects are distinct from other basic factors, such as IQ, and (2) associations between the distance effect, SNARC effect, and performance on a number line estimation task. As for whether the fractions SNARC and NLE performance would predict math achievement in our sample, we did not have strong *a priori* predictions due to the conflicting nature of relevant theory and past research. Theoretically, a stronger internal spatial-numerical representation (i.e., MNL) should be associated with higher mathematical achievement. Additionally, non-symbolic ratio comparison has been shown to predict university algebra scores (Matthews et al., 2016), and NLE performance has been associated with greater mathematical competence (Schneider et al., 2018). However, the SNARC effect with whole numbers has not been positively associated with math proficiency (e.g., Hoffmann et al., 2014; Cipora et al., 2016). In light of these inconsistent findings, we hypothesized that the slope of participants' fraction SNARC effects and NLE performance might uniquely account for variability in more domain-specific outcome measures, such as a formal test of fraction knowledge and a standardized measure of basic math skills, but would not predict algebra scores.

## METHODS AND MEASURES

### Participants and Procedure

One hundred and six undergraduate students were recruited for this study. However, no data was collected for one participant, as the session was disrupted shortly after the start. Thus, the final sample consisted of 105 adults, aged 18–43 (mean = 20.39 years,  $SD = 2.83$ ), who participated in this study for course credit. All components of the study were approved by the Institutional Review Board (IRB#2013-1346). Computerized experiments were programmed with E-prime 2.0.8.90a (Psychology Software Tools, Sharpsburg, PA, United States) on a Dell Optiplex 390 Desktop PC (3.1 GHz, 4 GB RAM) running Windows 7.0 64-bit operating system. Visual stimuli were presented on a Dell UltraSharp U2212H 21.5" flat-screen monitor at a resolution of  $1024 \times 768$  and a refresh rate of 60 Hz.

### Measures

The study session lasted approximately 1.5 h, during which time participants completed several measures, in following order:

### Fraction Comparison

Participants compared all 26 single-digit, irreducible fractions to the standard fraction  $\frac{1}{2}$ , indicating with a keyboard response if the fraction was larger or smaller than the standard. In an exact replication of Experiment 2 from Toomarian and Hubbard (2018b), each fraction appeared eight times, with response side counterbalanced across two blocks and two different run orders. A total of 10 practice trials preceded each block, which included visual feedback. A central fixation cross appeared for 600 ms, followed by a blank screen for 1000 ms and the target fraction for 3000 ms or until a response was detected. Fraction stimuli were approximately 1.8 cm wide and 2.7 cm tall ( $1.5^\circ \times 2.8^\circ$  visual angle). Left button presses corresponded to the 'd' key, and right button presses corresponded to the 'k' key on the QWERTY keyboard (distance = 8.5 cm).

Left hand median reaction times were subtracted from left hand median reaction times for each fraction magnitude for each participant. These differences in reaction times (dRT) were regressed on fraction magnitude, resulting in either a positive or negative sloping regression line for each participant (Lorch and Myers, 1990; Fias et al., 1996). Negative slopes indicate a classic SNARC effect (small magnitudes associated with the left, large with right), and positive slopes indicate the reverse. Data from this task yielded several outcome measures: an individual SNARC effect, individual distance effect, overall RT, and overall accuracy. It is important to note that this task is based on a direct magnitude comparison rather than the classic parity judgment primarily because fractions cannot be classified as even or odd.

### Number Line Estimation (NLE)

This computerized number-to-position task included both proper fractions on a 0–1 number line and improper fractions on a 0–5 number line (adapted from Torbeyns et al., 2015). Specifically, participants estimated the position on a number line that corresponded with the fraction displayed at the top of the screen. On the basis of these estimates, we calculated the percent absolute error (PAE) score for each participant ( $PAE = \frac{|\text{answer} - \text{correct answer}|}{\text{numerical range}}$ ). Thus, smaller PAE values indicate higher acuity for fractions.

### Fraction Knowledge Assessment (FKA)

This written assessment of fraction knowledge is comprised of items largely drawn from the TIMSS and NAEP (Matthews et al., 2016). Items were intended to assess both procedural (e.g., " $\frac{1}{10} + \frac{3}{5} = \underline{\quad}$ ") and conceptual (e.g., "How many fractions are possible fractions are between  $\frac{1}{4}$  and  $\frac{1}{2}$ ?" ) fraction knowledge. The assessment had a total possible score of 38 points; percentage correct was used as a quantitative measure of general fraction knowledge for each participant.

### Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)

This standardized assessment was used to quickly generate an estimate of IQ. Administration of two subtests—Vocabulary and Matrix Reasoning (MR)—yielded the Full Scale IQ 2 (FSIQ-2). Scores for Matrix Reasoning were also used as a measure of abstract problem solving, inductive reasoning and spatial reasoning.

## Placement Exams

Participants provided consent for the study team to obtain placement test scores from university administration. All students entering the University of Wisconsin system take a required series of math and English placement tests, comprised of Basic Mathematics, Algebra, Trigonometry, English, and Reading scores. Of particular theoretical interest are the Basic Math and Algebra scores, which have strong internal consistency (Cronbach's  $\alpha = 0.90$ ) and have been linked to non-symbolic ratio processing ability (Matthews et al., 2016). Scores are standardized on a scale ranging from 150 to 850 points.

## RESULTS

The accuracy threshold for inclusion was 80%, but all participants who completed the session exceeded this threshold. Missing data due to various technical issues (e.g., computer error, fire alarms) resulted in several participants without data for all of the measures conducted in a session. Additionally, placement test scores were unavailable for 19 participants. Thus, the following analyses describe results from slightly different samples, dependent on which measures were available for each participant. Sample sizes for each analysis are listed in **Table 1**, along with descriptive statistics. Diagnostic analyses revealed two influential points (as measured by Cook's  $d$ ). These outlier points reflected extreme but not implausible values, and removal of these two points did not meaningfully change the regression results. Thus, all possible data points were retained in the following models. SNARC effects were analyzed using regression analyses of repeated-measures data and  $t$ -tests against zero. This method has come to be favored over using an ANOVA as magnitudes can be analyzed continuously and accounts for between-subjects variability (for additional rationale on this approach, see Fias et al., 1996). This approach is particularly useful for investigations of individual differences, as it yields a SNARC slope for each participant which can then be used

in further analyses (e.g., correlations). Due to incongruous scaling of the measures, all reported beta values reflect standardized regression coefficients. Outcome measures were not standardized. There was no evidence of multicollinearity among the factors included in the model, as evidenced by variance inflation factors less than 10.

## Distance and SNARC Effects

As predicted, there was a significant group-level distance effect, both when average RTs were regressed on magnitude ( $\beta = -840.11$ ,  $F[1,11] = 105.8$ ,  $p < 0.001$ ) and when individual distance effects were tested against zero in a one-sample  $t$ -test ( $\beta = -912.85$ ,  $t[1,98] = -24.31$ ,  $p = 0.007$ ). Consistent with Toomarian and Hubbard (2018b), individual SNARC slopes were overall significantly less than zero ( $\beta = -75.57$ ,  $t[1,98] = -2.72$ ,  $p < 0.001$ ), indicating a group-level classic SNARC effect for fractions.

## Correlational Analyses

Simple bivariate correlations for all measures in the study are listed in **Table 2**. There was no correlation between the distance effect and SNARC effect ( $r = 0.05$ ,  $p = 0.622$ ). When accounting for the possible mediating role of RT, the correlation was still non-significant ( $p = 0.54$ ). The fractions SNARC was correlated with both acuity on the NLE task (PAE;  $r = 0.23$ ,  $p = 0.029$ ) and basic math ability (MBSC,  $r = -0.26$ ,  $p = 0.018$ ), meaning that increasingly negative SNARC slopes were associated with lower PAE scores (greater acuity) on the fractions NLE task and better basic math scores. Lower PAE was also associated with higher scores on the fractions task (FKA;  $r = -0.42$ ,  $p < 0.001$ ), higher accuracy on the fraction comparison task (ACC;  $r = -0.33$ ,  $p = 0.001$ ), basic math scores ( $r = -0.26$ ,  $p = 0.024$ ), and algebra scores (ALG;  $r = -0.26$ ,  $p = 0.023$ ).

## Predicting the SNARC Effect

To investigate our first research question of which factors predict the SNARC effect, we used linear regression to model the following equation:  $SNARC_i = \alpha + \beta_1 MR + \beta_2 Vocab + \beta_3 PAE + \beta_4 RT + \beta_5 ACC + \epsilon$  (see **Table 3**). The only significant factor in the specified model was performance on the number line estimation task. When holding all other factors constant, for every standard deviation increase in PAE (i.e., decreasing acuity), the SNARC slope is expected to increase by 82.88 ( $t = 2.76$ ,  $p = 0.007$ ), resulting in an increasingly positive slope. In other words, acuity for a physical number line task—as measured by PAE—uniquely predicts the degree to which participants activate holistic fraction magnitudes on their (implicit) mental number line. Indices of general intelligence, RTs, and accuracy did not meaningfully influence the fraction SNARC. This provides some validation that the fraction SNARC effect is a valuable measurement of internal SNAs and is distinct from other measures of task performance. However, this model predicted relatively little variance in SNARC slopes, suggesting that other factors (not measured in this investigation) have greater influence on the variability in individuals' SNARC effects.

**TABLE 1 |** Descriptive statistics.

Measure	<i>n</i>	Mean (SD)
Fraction comparison		
Reaction time (RT)	99	749.44 (137.24)
Accuracy (ACC)	99	0.96 (0.02)
SNARC slope (SNARC)	99	-75.57 (276.32)
Distance Effect slope (DIST)	99	-912.85 (373.67)
Fraction Knowledge Assessment % (FKA)	100	84.11 (10.28)
Number Line Estimation (PAE)	94	6.89 (2.75)
Algebra Exam (ALG)	86	585.00 (101.80)
Basic Math Exam (MBSC)	86	629.19 (104.87)
WASI- Full-Scale IQ (FSIQ)	102	104.33 (10.50)
Matrix Reasoning (MR)	102	49.81 (8.26)
Vocabulary (VOCAB)	102	55.36 (6.57)

Descriptions of each measure include the abbreviation used in subsequent analyses. Reaction time measured in milliseconds. SNARC, spatial-numerical association of response codes.



**TABLE 2 |** Bivariate correlations.

	FKA	SNARC	FSIQ	RT	ACC	DIST	PAE	MBSC	ALG	MR
SNARC	-0.15	1								
FSIQ	0.26**	-0.09	1							
RT	-0.14	-0.01	0.06	1						
ACC	0.26**	0.01	0.11	0.20*	1					
DIST	0.09	0.05	0.03	-0.69***	-0.27**	1				
PAE	-0.42***	0.23*	0.02	0.19	-0.33**	-0.06	1			
MBSC	0.43***	-0.26*	0.36***	-0.02	0.09	-0.06	-0.26*	1		
ALG	0.33**	-0.17	0.33**	-0.18	0.15	0.09	-0.26*	0.70***	1	
MR	0.26**	-0.13	0.86***	-0.01	0.10	0.12	-0.07	0.29**	0.34**	1
Vocab	0.18	-0.01	0.76***	0.11	0.07	-0.04	0.10	0.33**	0.18	0.36***

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ .

**TABLE 3 |** Regression analysis for variables predicting SNARC effect slope.

Variable	$\beta$	SE
Intercept	-74.51	28.31
WASI- MR	-15.71	32.42
WASI - Vocab	-8.85	30.43
Number Line Est. (PAE)	81.12*	31.10
RT	-10.95	29.95
ACC	52.00	30.77
R-squared	0.086	
Adjusted R-Squared	0.032	

\* $p < 0.05$ .  $\beta$  represents standardized regression coefficients.  $n = 90$ .

## Contributions to Fraction Knowledge

Next, we aimed to test the unique contributions of SNARC slopes and PAE to procedural and conceptual fraction knowledge, as measured by the FKA. To do this, we conducted a three-step hierarchical regression analysis that introduced SNARC and then PAE to the reduced model containing other basic cognitive factors that could influence FKA scores (see **Table 4**). Because participants with any missing values for SNARC, PAE or FKA were excluded from analysis, 88 participants were retained for this analysis. Step 1 included only mean RT, mean accuracy, and full scale IQ, which together accounted for 14% of the variance in FKA scores ( $F[3,84] = 5.45$ ,  $p = 0.002$ ). All of these factors on their own predicted FKA scores. When SNARC slopes were added in Step 2, only an additional 1% of variance in FKA scores was accounted for, and it was not significantly improved from the reduced model ( $F[1,83] = 3.003$ ,  $p = 0.09$ ). In the third step, PAE from the NLE task was added to the model, which increased the amount of explained variance in FKA scores to 23%, a significant improvement in model specification ( $F[1,82] = 9.35$ ,  $p = 0.003$ ) compared to the model in Step 2.

Notably, there was no evidence of multicollinearity among the factors included in the model, as evidenced by relatively small variance inflation factors (SNARC slope = 1.16; PAE = 1.29; RT = 1.17, ACC = 1.35, IQ = 1.03). When all other basic cognitive factors and the SNARC are controlled for, FKA scores decrease by 0.03 points for each standard deviation increase in PAE for the fractions number line task. To summarize, scores on a

fraction test were significantly predicted by an explicit number line estimation task but not by an implicit measure of SNAs for fractions, contrary to our initial hypothesis.

## Contributions to Basic Math Skills

To investigate the relative contributions of implicit and explicit processing of SNAs to basic math skills, we conducted another three-step hierarchical regression analysis, with progressive introduction of the SNARC effect and then PAE score as predictors. The first model contained the same initial predictors as the previous model for FKA scores, namely RT, ACC, and FSIQ (see **Table 5**). Because participants with any missing values for SNARC, PAE or MBSC were excluded from analysis, 73 participants were retained for this analysis.

This first regression model explained 7% of the variance in scores for basic math skills ( $F[3,69] = 2.78$ ,  $p = 0.05$ ). In this reduced sample, only FSIQ predicted scores on MBSC, meaning that when holding all other factors constant, each standard deviation increase in FSIQ is associated with a 38.19 point increase in MBSC score. The addition of SNARC slopes explained 1% more variance, though according to a partial F-test, this model was not a significant improvement ( $F[1,68] = 1.58$ ,  $p = 0.21$ ). The last step—adding in PAE—resulted in a slightly better model and explained an additional 3% of variance in MBSC scores ( $F[1,67] = 4.13$ ,  $p = 0.05$ ). For each standard deviation increase in PAE (indicating reduced acuity), MBSC scores decrease by 26.69 points, controlling for changes in ACC, RT, FSIQ, and SNARC.

## Contributions to Algebraic Knowledge

The last outcome measure we tested was score on a standardized algebra exam. This outcome measure was motivated by findings that college students' non-symbolic ratio judgments significantly predicted algebra placement exam scores (Matthews et al., 2016). To test whether either the SNARC or PAE predicted algebra scores, we conducted another three-step hierarchical regression analysis to investigate the relative contributions of implicit and explicit measures of SNAs to ALG. These models followed the same structure as the previous two hierarchical regression models, with basic cognitive factors in the initial model, followed by progressive introduction SNARC and PAE



**TABLE 4 |** Hierarchical regression analysis for variables predicting FKA score.

Predictor variable	Regression 1		Regression 2		Regression 3	
	$\beta$	SE	$\beta$	SE	$\beta$	SE
RT	−0.02	0.01	−0.02*	0.01	−0.01	0.01
ACC	0.03**	0.01	0.03**	0.01	0.02	0.01
FSIQ	0.02*	0.01	0.02*	0.01	0.03*	0.01
SNARC			−0.02	0.01	−0.01	0.01
Number Line Est. (PAE)					−0.03**	0.01
$R^2$		0.13		0.15		0.23
$\Delta R^2$				0.02		0.07**

\* $p < 0.05$ , \*\* $p < 0.01$ , all reported  $R^2$  are adjusted.  $n = 88$ .

**TABLE 5 |** Hierarchical regression analysis for variables predicting basic math score.

Predictor Variable	Regression 1		Regression 2		Regression 3	
	$\beta$	SE	$\beta$	SE	$\beta$	SE
RT	−5.10	11.87	−7.29	11.95	−1.98	11.97
ACC	2.01	11.74	4.96	11.92	−4.79	12.61
FSIQ	38.19**	13.62	36.35**	13.64	38.31**	13.37
SNARC			−15.52	12.35	−8.72	12.53
Number Line Est. (PAE)					−26.69*	13.14
$R^2$		0.07		0.08		0.11
$\Delta R^2$				0.01		0.03*

\* $p < 0.05$ , \*\* $p < 0.01$ , all reported  $R^2$  are adjusted.  $n = 73$ .

**TABLE 6 |** Hierarchical regression analysis for variables predicting algebra scores.

Predictor Variable	Regression 1		Regression 2		Regression 3	
	$\beta$	SE	$\beta$	SE	$\beta$	SE
RT	−28.35	10.49**	−29.19	10.66**	−25.57	10.82*
ACC	17.02	10.38	18.15	10.64	11.50	11.39
FSIQ	23.59	12.04	22.88	12.17	24.22	12.08*
SNARC			−5.95	11.02	−1.32	11.32
Number Line Est. (PAE)					−18.19	11.87
$R^2$		0.12		0.11		0.13
$\Delta R^2$				−0.01		0.02

\* $p < 0.05$ , \*\* $p < 0.01$ , all reported  $R^2$  are adjusted.  $n = 73$ .

score (Table 6). Due to incomplete cases, 73 participants were retained for analysis.

In the initial model, only RT was a significant predictor of algebra test scores ( $p = 0.008$ ), and 12% of the variance in ALG was explained by the model. When SNARC was introduced, the model actually explained less variance, when the number of factors was considered (adj- $R^2 = 0.11$ ). Adding PAE to the model explained an additional 1% of variance from the first model, though neither of the subsequent models were any better than the first (1 vs. 2:  $F[1,68] = 0.29$ ,  $p = 0.59$ ; 2 vs. 3:  $F[1,67] = 2.35$ ,  $p = 0.13$ ), indicating that neither implicit nor explicit measures of SNAs have predictive power over algebra test scores. In the final model, only RT and FSIQ significantly predicted ALG. Thus,

while holding all other variables in the final regression constant, ALG scores increase by 25.57 points for every standard deviation decrease in RT; they increase by 24.22 points for every standard deviation increase in FSIQ.

## Mediation Analyses

Despite the extensive planned analyses, it is unclear whether SNARC slopes and PAE scores contribute uniquely to our outcomes of interest, specifically FKA and MBSC scores. We employed mediated path analyses to determine whether acuity on the NLE task—as measured by PAE—mediated the relationship between the SNARC and our two outcome measures of interest. We did not have reason to believe that there was any mediation

in the case of ALG scores, since neither measure was predictive of ALG scores in prior analyses. Additionally, while the independent variable predicting the dependent variable is often regarded as a necessary condition for conducting mediation analyses (Baron and Kenny, 1986), recent guidelines have supported mediation analysis without such a relationship in certain cases (Shrout and Bolger, 2002). For instance, in cases when theory would predict such a relationship and sample sizes are relatively small, mediation analysis may be conducted with bootstrapped confidence intervals. Thus, although SNARC did not predict FKA scores, we proceeded with mediated path analysis nonetheless. To test whether PAE mediates the relationship between SNARC and our two dependent measures (FKA and MBSC), we conducted path analysis with mediation using the 'lavaan' package in R (Rosseel, 2012). Variables are unstandardized. We used the full information maximum-likelihood imputation approach for missing values.

In Model A (Figure 1), the only direct effect was between NLE and FKA scores; adjusting for SNARC slopes, every 1-unit increase in PAE is associated with a decrease of  $b = 0.568$  ( $SE = 0.16$ ,  $p < 0.001$ ) in FKA score. There was no indirect effect, and thus no evidence of full mediation  $ab = -0.001$  ( $SE = 0.0008$ ,  $p = 0.204$ ). A bias-corrected bootstrapped 95% confidence interval based on 10,000 samples included zero  $[-0.003, 0.0001]$ , confirming that there is no evidence of mediation in this model.

In Model B, we tested for mediation between SNARC and MBSC score. Independent of PAE, a one-unit increase in SNARC slope is associated with 0.107 decrease in MBSC score ( $SE = 0.044$ ,  $p = 0.014$ ). Every unit increase in SNARC slope is associated with an  $a = 0.003$  ( $SE = 0.001$ ,  $p = 0.028$ ) increase in PAE on the NLE task. Adjusting for SNARC slopes, every 1-unit increase in PAE is associated with a decrease of  $b = 9.983$  ( $SE = 4.400$ ,  $p = 0.023$ ) in MBSC score. There was no indirect effect, and thus no evidence that PAE score mediated this association  $ab = -0.026$  ( $SE = 0.019$ ,  $p = 0.184$ ). A bias-corrected bootstrapped 95% confidence interval based on 10,000 samples included zero  $[-0.077, 0.0002]$ , confirming that there is no evidence of full mediation in this model. However, there was a significant total effect for the model ( $SE = 0.044$ ,  $p = 0.015$ ), indicating that the

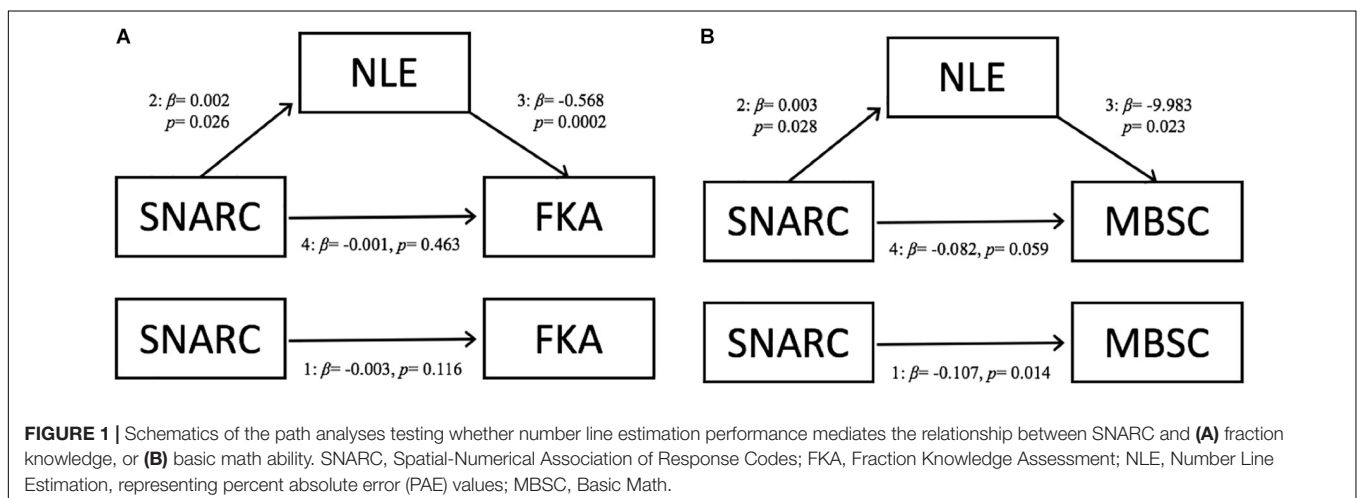
model fit the data well and is evidence that PAE may at least partially mediate the relationship between SNARC and MBSC.

## DISCUSSION

In this study, we investigated the relationship between implicit and explicit measures of SNAs, including the link to formal math abilities. First, we successfully replicated our previous work demonstrating that a classic SNARC for fraction magnitudes emerges at the group-level (Toomarian and Hubbard, 2018b) and for the majority of adult individuals. This replication in a separate, larger sample of adults supports the assertion that people can and do represent fractions holistically under appropriate task constraints.

We then moved past group level effects to investigate a second question: which factors influence individual differences in participants' SNARC effects. Performance on a number line estimation task, which included whole numbers and fractions, was uniquely predictive of individual SNARC slopes. Importantly, this relationship emerged even while controlling for factors such as response time, overall accuracy, and two IQ subtests. That accuracy and RT in the comparison task were not associated with SNARC slopes indicates that the SNARC is measuring a unique, spatial ability that cannot be accounted for by basic processing speed or ability to do the task. These results are theoretically supported by the MNL hypothesis; if the SNARC is a measure of reliance on a right-to-left spatially oriented MNL, greater reliance on this internal number line (evidenced by more negative SNARC slopes) should be related to acuity on a similarly oriented, external number line task. However, Schneider et al. (2009) found no relationship between NLE performance and the parity SNARC in kids, thereby challenging this interpretation of the results. Instead, they argue that the internal and external number line cannot be equated, at least early in development.

Our results indicate that NLE has greater predictive power than the SNARC for multiple outcome measures, which suggests some degree of dissociation between these two measures. One explanation for this dissociation may be that the fractions



SNARC, by nature of being more implicit than the NLE task, has a weaker effect and may not have much influence to exert on explicit outcome measures. This is in contrast to the NLE task, which has both theoretical (e.g., Siegler et al., 2011) and empirical (e.g., Thompson and Siegler, 2010; Gunderson et al., 2012; Resnick et al., 2016; Ye et al., 2016) support for its role in fractions learning and math proficiency. A recent study demonstrated that number line training but not area model training improved performance on an untrained fraction magnitude comparison task, highlighting the utility of an external spatial-numerical representation (Hamdan and Gunderson, 2017).

In this study, there was no evidence of a correlation between the distance effect and SNARC effect. Previous studies with whole numbers have yielded mixed evidence on the relationship between the distance and SNARC effects; Viarouge et al. (2014) found a correlation between these measures, while Gibson and Maurer (2016) did not. Interestingly, Schneider et al. (2009) found a significant correlation in one experiment, but not in a subsequent experiment.<sup>1</sup> While both effects are often taken as evidence supporting the MNL hypothesis, there is a key difference between the two effects: only the SNARC effect reflects a directional/spatialized association. With this difference in mind, it is not difficult to imagine that these effects might dissociate within subjects, particularly for stimuli such as common fractions, for which the cognitive processing mechanisms are still not well understood.

Lastly, neither the fractions SNARC nor PAE predicted algebra placement exam scores, despite PAE being a significant predictor of fraction knowledge and basic math skills. This suggests that more implicit processing of spatial-numerical representation may not be as readily recruited during higher-order mathematical concepts, but rather may serve as a foundation for thinking about simpler problems involving rational magnitudes. This would cohere well with the recent finding that the ability to place decimals, but not fractions, on number lines was one of the best predictors of algebra performance (DeWolf et al., 2015).

## Limitations

Here we would like to note several aspects of the current research that may limit the interpretability of the results. First, as previously mentioned, the sample size was moderately reduced for each analysis due to missing data points across various measures. This issue was perhaps most significant for the hierarchical regressions with MBSC and ALG as the dependent variables, since the placement tests were the variables for which there were the most missing data points. While this reduction affected the degrees of freedom, decreased the adjusted R-squared, and increased the possible influence of outliers, it is important to note that the total  $n$  never dipped below the

number required for a medium effect size and there were no marginal effects.

Additionally, recent simulation work on detecting reliable SNARC effects with various sample sizes, stimulus repetitions, and effects has provided guidelines for obtaining results of moderate effect (Cipora and Wood, 2017). Specifically, studies are recommended to test a minimum of 20 participants and with twenty repetitions per stimulus. While our sample size exceeds this minimum requirement, there are only eight repetitions per stimulus in the task from which we draw our individual SNARC slopes. That said, our stimulus set contains four times the number of individual numerical stimuli as classic SNARC paradigms (24 vs. 8), thus offsetting the reduction in the number of trials per stimulus. Thus, the overall experiment time would be unreasonably long if we were to collect twenty observations per stimulus per condition and would thus compromise the integrity of the data. Furthermore, because this recommendation stems from the desire to control for intra-individual variability, we argue that our wide range of fraction magnitudes in fact serves a similar purpose; by increasing the number of points on the MNL to which participants are asked to respond, we are effectively controlling for this variability in an analogous fashion.

## CONCLUSION

In this study, we investigated how individual spatial representations of fractions relate to explicit fraction knowledge and two other formal measures of math achievement. We observed significant group-level SNARC and distance effects based on overall fraction magnitude, with notable individual variability. Performance for the number line estimation task was correlated with SNARC slopes and predicted significant variance in SNARC slopes even when accounting for factors such as overall accuracy and matrix reasoning ability. Multi-step regressions revealed that NLE performance was a significant predictor of fraction test scores and basic math skills but the SNARC was not, indicating that working with an explicit number line may be a stronger predictor of domain-specific and domain-general math abilities than more implicit number line processing of fractions. Neither individual SNARC effects nor NLE performance were significant predictors of algebra scores. This suggests that the MNL may not be as readily recruited during higher-order mathematical concepts, but rather may be a foundation for thinking about simpler problems involving rational magnitudes.

The current study informs our understanding of the relative contributions of more implicit (SNARC) and explicit (NLE) processing of fractions, but it is still unknown whether these relations are consistent from childhood to adulthood. Developmental studies—particularly with continuous age data—are necessary to better understand how spatial and numerical conceptions influence mathematical thinking. Future studies should investigate this relationship with (1) a larger, more educationally-diverse sample, and (2) additional spatial tasks as covariates.

<sup>1</sup>Beyond just significance testing, these studies also found markedly different correlation coefficients for the relationship between SNARC and distance effect: Viarouge et al. (2014):  $r = 0.52$ ; Schneider et al. (2009):  $r = 0.25$  (Experiment 1) and  $r = -0.03$  (Experiment 2); Gibson and Maurer (2016):  $r = -0.06$ ; the current study:  $r = 0.05$ .

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the University of Wisconsin-Madison Institutional Review Board (IRB#2013-1346) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Educational/Social Behavioral Sciences (Ed/SBS) IRB at UW-Madison.

## AUTHOR CONTRIBUTIONS

ET and EH conceptualized and designed the study. ET collected the data. ET and RM analyzed the data. ET wrote the first draft of the manuscript. All authors revised, read, and approved the submitted version of manuscript.

## REFERENCES

- Abrahamse, E., Van Dijck, J.-P., and Fias, W. (2016). How does working memory enable number-induced spatial biases? *Front. Psychol.* 7:977. doi: 10.3389/fpsyg.2016.00977
- Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social 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
- Booth, J. L., and Newton, K. J. (2012). Fractions: could they really be the gatekeeper's doorman? *Contemp. Educ. Psychol.* 37, 247–253. doi: 10.1016/j.cedpsych.2012.07.001
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189
- 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
- Cipora, K., Hohol, M., Nuerk, H.-C., Willmes, K., Brożek, B., Kucharzyk, B., et al. (2016). Professional mathematicians differ from controls in their spatial-numerical associations. *Psychol. Res.* 80, 710–726. doi: 10.1007/s00426-015-0677-6
- Cipora, K., and Nuerk, H. (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
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- DeWolf, M., Bassok, M., and Holyoak, K. J. (2015). From rational numbers to algebra: separable contributions of decimal magnitude and relational understanding of fractions. *J. Exp. Child Psychol.* 133, 72–84. doi: 10.1016/j.jecp.2015.01.013
- 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
- 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
- Fumarola, A., Prpic, V., Fornasier, D., Sartoretto, F., Agostini, T., and Umiltà, C. (2016). The spatial representation of angles. *Perception* 45, 1320–1330. doi: 10.1177/0301006616661915
- Galton, F. (1880). Visualised numerals. *Nature* 21, 494–495.

## FUNDING

This research was supported by the National Science Foundation (DGE-1256259). The opinions expressed are those of the authors and do not represent the views of the National Science Foundation.

## ACKNOWLEDGMENTS

We would like to thank Rebecca Liu, Nina Vakil, and Carolyn Heal for assistance with data collection; Percival Matthews, Martina Rau, Martha Alibali, Andreas Obersteiner and members of the Educational Neuroscience Lab for helpful comments on the manuscript; and all of the participants who generously gave their time to this study. This work first appeared as one chapter of ET's dissertation.

- Gibson, L. C., and Maurer, D. (2016). Development of SNARC and distance effects and their relation to mathematical and visuospatial abilities. *J. Exp. Child Psychol.* 150, 301–313. doi: 10.1016/j.jecp.2016.05.009
- Gilligan, K. A., Flouri, E., and Farran, E. K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *J. Exp. Child Psychol.* 163, 107–125. doi: 10.1016/j.jecp.2017.04.016
- Gunderson, E. A., Ramirez, G., Beilock, S. L., and Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Dev. Psychol.* 48, 1229–1241. doi: 10.1037/a0027433
- Hamdan, N., and Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: evidence from a fraction intervention. *Dev. Psychol.* 53, 587–596. doi: 10.1037/dev0000252
- Hoffmann, D., Mussolin, C., Martin, R., and Schiltz, C. (2014). The impact of mathematical proficiency on the number-space association. *PLoS One* 9:e85048. doi: 10.1371/journal.pone.0085048
- Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448. doi: 10.1038/nrn1684
- Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2009). Numerical and spatial intuitions: a role for posterior parietal cortex? *Cogn. Biol. Evol. Dev. Perspect. Mind Brain Behav.* 11, 221–246.
- Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., and Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *J. Exp. Child Psychol.* 116, 45–58. doi: 10.1016/j.jecp.2013.02.001
- Lewis, M. R., Matthews, P. G., and Hubbard, E. M. (2015). "Neurocognitive architectures and the nonsymbolic foundations of fractions understanding," in *Development of Mathematical Cognition-Neural Substrates and Genetic Influences*, eds K. M. Koepke, D. Berch, and D. Geary (Amsterdam: Elsevier).
- Lorch, R. F., and Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 149–157. doi: 10.1037/0278-7393.16.1.149
- Matthews, P. G., Lewis, M. R., and Hubbard, E. M. (2016). Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychol. Sci.* 27, 191–202. doi: 10.1177/0956797615617799
- Mazzocco, M. M. M., Myers, G. F., Lewis, K. E., Hanich, L. B., and Murphy, M. M. (2013). Limited knowledge of fraction representations differentiates middle school students with mathematics learning disability (dyscalculia) versus low mathematics achievement. *J. Exp. Child Psychol.* 115, 371–387. doi: 10.1016/j.jecp.2013.01.005
- Mix, K. S., and Cheng, Y. L. (2012). The relation between space and math. developmental and educational implications. *Adv. Child Dev. Behav.* 42, 197–243. doi: 10.1016/B978-0-12-394388-0.00006-X
- Möhring, W., Newcombe, N. S., and Frick, A. (2015). The relation between spatial thinking and proportional reasoning in preschoolers. *J. Exp. Child Psychol.* 132, 213–220. doi: 10.1016/j.jecp.2015.01.005



- Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0
- Muldoon, K., Towse, J., Simms, V., Perra, O., and Menzies, V. (2013). A longitudinal analysis of estimation, counting skills, and mathematical ability across the first school year. *Dev. Psychol.* 49, 250–257. doi: 10.1037/a0028240
- Nuerk, H.-C., Patro, K., Cress, U., Schild, U., Friedrich, C. K., and Göbel, S. M. (2015). How space-number associations may be created in preliterate children: six distinct mechanisms. *Front. Psychol.* 6:215. doi: 10.3389/fpsyg.2015.00215
- Nuerk, H. C., Wood, G., and Willmes, K. (2005). The universal SNARC effect: the association between number magnitude and space is amodal. *Exp. Psychol.* 52, 187–194. doi: 10.1027/1618-3169.52.3.187
- Proctor, R. W., and Xiong, A. (2015). Polarity correspondence as a general compatibility principle. *Curr. Dir. Psychol. Sci.* 24, 446–451. doi: 10.1177/0963721415607305
- Prpic, V., Soranzo, A., Santoro, I., Fantoni, C., Galmonte, A., Agostini, T., et al. (2018). SNARC-like compatibility effects for physical and phenomenal magnitudes: a study on visual illusions. *Psychol. Res.* doi: 10.1007/s00426-018-1125-1 [Epub ahead of print].
- 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.
- Ren, P., Nicholls, M. E. R., Ma, Y., and Chen, L. (2011). Size matters: non-numerical magnitude affects the spatial coding of response. *PLoS One* 6:e23553. doi: 10.1371/journal.pone.0023553
- Resnick, H., Jordan, N. C., Hansen, N., Rajan, V., Rodrigues, J., Siegler, R. S., et al. (2016). Developmental growth trajectories in understanding of fraction magnitude from fourth through sixth grade. *Dev. Psychol.* 12, 1–12. doi: 10.1037/dev0000102
- Restle, F. (1970). Speed of adding and comparing numbers. *J. Exp. Psychol.* 83, 274–278. doi: 10.1037/h0028573
- Rosseel, Y. (2012). Ivaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.3389/fpsyg.2014.01521
- Schneider, M., Grabner, R. H., and Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *J. Educ. Psychol.* 101, 359–372. doi: 10.1037/a0013840
- 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.* 89, 1467–1484. doi: 10.1111/cdev.13068
- Shea, D. L., Lubinski, D., and Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. *J. Educ. Psychol.* 93, 604–614. doi: 10.1037/0022-0663.93.3.604
- Shrout, P. E., and Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol. Methods* 7, 422–445. doi: 10.1037/1082-989X.7.4.422
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Siegler, R. S., and Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Dev. Psychol.* 49, 1994–2004. doi: 10.1037/a0031200
- 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
- Siegler, R. S., Thompson, C. A., and Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cogn. Psychol.* 62, 273–296. doi: 10.1016/j.cogpsych.2011.03.001
- Simms, V., Clayton, S., Cragg, L., Gilmore, C., and Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: the role of visuomotor integration and visuospatial skills. *J. Exp. Child Psychol.* 145, 22–33. doi: 10.1016/j.jecp.2015.12.004
- Thompson, C. A., and Siegler, R. S. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychol. Sci.* 21, 1274–1281. doi: 10.1177/0956797610378309
- Toomarian, E. Y., and Hubbard, E. M. (2018a). On the genesis of spatial-numerical associations: evolutionary and cultural factors co-construct the mental number line. *Neurosci. Biobehav. Rev.* 90, 184–199. doi: 10.1016/j.neubiorev.2018.04.010
- Toomarian, E. Y., and Hubbard, E. M. (2018b). The fractions snarc revisited: processing fractions on a consistent mental number line. *Q. J. Exp. Psychol.* 71, 1761–1770. doi: 10.1080/17470218.2017.1350867
- Torbeyns, J., Schneider, M., Xin, Z., and Siegler, R. S. (2015). Bridging the gap: fraction understanding is central to mathematics achievement in students from three different continents. *Learn. Instr.* 37, 5–13. doi: 10.1016/j.learninstruc.2014.03.002
- Viarouge, A., Hubbard, E. M., and McCandliss, B. D. (2014). The cognitive mechanisms of the SNARC effect: an individual differences approach. *PLoS One* 9:e95756. doi: 10.1371/journal.pone.0095756
- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127
- Whyte, J. C., and Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Dev. Psychol.* 44, 588–596. doi: 10.1037/0012-1649.44.2.588
- Wood, G., Nuerk, H.-C., and Willmes, K. (2006). Crossed hands and the SNARC effect: a failure to replicate Dehaene, Bossini and Giraux (1993). *Cortex* 42, 1069–1079. doi: 10.1016/S0010-9452(08)70219-3
- Ye, A., Resnick, I., Hansen, N., Rodrigues, J., Rinne, L., and Jordan, N. C. (2016). Pathways to fraction learning: numerical abilities mediate the relation between early cognitive competencies and later fraction knowledge. *J. Exp. Child Psychol.* 152, 242–263. doi: 10.1016/j.jecp.2016.08.001

**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 Toomarian, Meng and Hubbard. 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.



# Multisensory Interactive Technologies for Primary Education: From Science to Technology

Gualtiero Volpe<sup>1\*</sup> and Monica Gori<sup>2</sup>

<sup>1</sup>Casa Paganini-InfoMus, DIBRIS, University of Genoa, Genoa, Italy, <sup>2</sup>U-Vip Unit, Istituto Italiano di Tecnologia, Genoa, Italy

## OPEN ACCESS

### Edited by:

Claudio Longobardi,  
University of Turin, Italy

### Reviewed by:

Benjamin A. Rowland,  
Wake Forest University,  
United States  
Sharlene D. Newman,  
Indiana University Bloomington,  
United States

### \*Correspondence:

Gualtiero Volpe  
gualtiero.volpe@unige.it

### Specialty section:

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

**Received:** 04 December 2018

**Accepted:** 25 April 2019

**Published:** 28 June 2019

### Citation:

Volpe G and Gori M (2019)  
Multisensory Interactive  
Technologies for Primary Education:  
From Science to Technology.  
Front. Psychol. 10:1076.  
doi: 10.3389/fpsyg.2019.01076

While technology is increasingly used in the classroom, we observe at the same time that making teachers and students accept it is more difficult than expected. In this work, we focus on multisensory technologies and we argue that the intersection between current challenges in pedagogical practices and recent scientific evidence opens novel opportunities for these technologies to bring a significant benefit to the learning process. In our view, multisensory technologies are ideal for effectively supporting an embodied and enactive pedagogical approach exploiting the best-suited sensory modality to teach a concept at school. This represents a great opportunity for designing technologies, which are both grounded on robust scientific evidence and tailored to the actual needs of teachers and students. Based on our experience in technology-enhanced learning projects, we propose six golden rules we deem important for catching this opportunity and fully exploiting it.

**Keywords:** multisensory technologies, education, integration of sensory modalities, enaction, inclusion

## INTRODUCTION

Multisensory education is conceived as an instructional method using visual, auditory, kinesthetic, and tactile ways to educate students (Joshi et al., 2002). There has been a longstanding interest in how learning can be supported by representations engaging multiple modalities. For example, the Montessori education tradition makes use of artifacts such as sandpaper letters children trace with their fingers to develop the physical skill of learning to write. Papert (1980) discussed the idea of body-syntonic learning – projecting an experiential understanding of how bodies move – into learning about geometry. Moreno and Mayer (1999) explored the cognitive impact of multimodal learning material in reducing cognitive load by representing information in more than one modality.

Technology entered the classroom many years ago. It can be considered as a medium for inquiry, communication, construction, and expression (Bruce and Levin, 1997). Early technological interventions consisted of endowing classrooms with devices such as overhead projectors, cassette players, and simple calculators. These devices were intended to support the traditional learning and teaching paradigms and usually did not enable direct interaction of students with technology. More recently, a broad palette of technological tools became available, including technologies for computer-assisted instruction, i.e., the use of computers for tutorials or simulation activities offered in substitution or as a supplement to teacher-directed instruction (Hicks and Holden, 2007), and for computer-based instruction, i.e., the use of computers in the delivery of instruction (Kulik, 1983). These technologies exploit devices, such as interactive whiteboards, laptops,

smartphones, and tablets, which are mainly conceived to convey visual information and are not intended for embodied interaction. Walling (2014) argued that tablet computers are toolboxes for learner engagement and suggested that the transition to using tablet computers in education is a natural process for teenagers. For example, there are multiple applications for tablets for learning mathematics<sup>1</sup>. Falloon (2013) reviewed 45 apps selected by an experienced teacher. Of them, 27 were considered educational apps, which focused on a broad variety of topics, including numeracy skills, reinforcing spelling, acquiring new vocabulary, and improving phonetics. Nevertheless, these solutions usually rely on the ability to see digital content rather than physically interacting with it. At the same time, novel technological developments also enabled the use of multiple sensory channels, including the visual, auditory, and tactile ones. This technology has been defined as *multisensory technology*. Technological advances and increased availability of affordable devices (e.g., Kinect, Oculus Rift, and HTC Vive) allowed a fast adoption of multisensory technology in many areas (e.g., entertainment, games and exergames, and assistive technologies). Its introduction in the classroom, however, is still somewhat limited. Early works addressed the use of virtual reality in educational software for either enabling full immersion in virtual environments or accentuating specific sensory information (Raskind et al., 2005). Nowadays, technologies such as augmented reality (e.g., see Santos et al., 2014) and serious games (e.g., see Connolly et al., 2012) play a relevant role in many educational contexts, both in science and in the humanities. Multisensory technologies enabling embodied interaction were used, for example, to support teaching in computer programming (e.g., Katai and Toth, 2010; Katai, 2011), music (e.g., Varni et al., 2013), and dance (e.g., Rizzo et al., 2018). Baud-Bovy and Balzarotti (2017) reviewed recent research on force-feedback devices in educational settings, with a particular focus on primary school teaching. Less traditional tools were also exploited, including Job Access With Speech (JAWS) and Submersible Audible Light Sensor (SALS). JAWS is a computer screen reader program allowing blind and visually impaired users to read the screen. SALS is a glass wand with an embedded light sensor, enabling the measuring of color intensity changes. These tools were used in a science camp for visually impaired students (Supalo et al., 2011), but they could also be modified and adapted for general inclusion in multisensory education. Despite such initiatives and the growing interest in these tools, most often the introduction of multisensory technologies in the learning environment has been exploratory, piecemeal, or *ad hoc*, focusing on understanding the potentials of the different modalities, rather than taking a combined multisensory focus.

Stakeholders consider the adoption of a technology-mediated pedagogical approach as a must, and the quest for innovation heavily drives choices. Attempts at integrating technology in the classroom, however, do not often take into account the pedagogical needs and paradigms. Teachers and students are not involved in the innovation process, and development of

technologies does not follow a proper evidence-based iterative design approach. The risk is that technology can be rejected. For example, Groff and Mouza (2008) presented a literature review on the challenges associated with the effective integration of technology in the classroom. More recently, Johnson et al. (2016) discussed common challenges educators face when attempting to introduce technology at school. Philip (2017) described the difficulties that were experienced in a project relying on novel mobile technologies in the classroom.

In this article, we argue that the intersection between current challenges in pedagogical practices and recent scientific evidence opens novel opportunities for acceptance of technology as a tool for education, and those multisensory technologies can specifically bring a significant benefit to the teaching and learning process.

## SCIENTIFIC EVIDENCE

The combination and the integration of multiple unimodal units are crucial to optimize our everyday interaction with the environment (Ernst and Bulthoff, 2004). Sensory combination allows us to maximize information delivered by different sensory modalities without these modalities being necessarily fused, while sensory integration enables reducing the variance in the sensory estimate to increase its reliability (Ernst and Bulthoff, 2004). In particular, sensory combination occurs when different environmental properties of the same object are estimated by means of different sensory modalities. Contrarily, sensory integration occurs when the same environmental property is estimated by different sensory modalities (Ernst and Bulthoff, 2004). Many recent studies show that our brain is able to integrate unisensory signals in a statistically optimal fashion as predicted by a Bayesian model, weighting each sense according to its reliability (Clarke and Yuille, 1990; Ghahramani et al., 1997; Ernst and Banks, 2002; Alais and Burr, 2004; Landy et al., 2011). This model has been useful to predict the multisensory integration behavior of adults across different sensory modalities in an optimal or near-optimal fashion (Ernst and Banks, 2002; Alais and Burr, 2004; Landy et al., 2011). There is also firm neurophysiological evidence for multisensory integration. Studies in cats have demonstrated that the midbrain structure superior colliculus (SC) is involved in integrating information between modalities and in initiating and controlling localization and orientation of motor responses (Stein and Meredith, 1993). This structure is highly sensitive to input from the association cortex, and emergence of multisensory integration critically depends on cross-modal experiences that alter the underlying neural circuit (Stein et al., 2014). Moreover, cortical deactivation impairs integration of multisensory signals (Jiang et al., 2002, 2007; Rowland et al., 2014). Studies in monkeys explored multisensory decision making and underlying neurophysiology by considering visual and vestibular integration (Gu et al., 2008). Similar effects were also observed in rodents (Raposo et al., 2012, 2014; Sheppard et al., 2013).

The role of sensory modalities in child development has been the subject of relevant research in developmental psychology,

<sup>1</sup>See e.g., <http://www.pcadvisor.co.uk/feature/software/best-maths-apps-for-children-3380559/>

psychophysics, and neuroscience. On the one side, scientific results show that young infants seem to be able to match sensory information and benefit from the presence of congruent sensory signals (Lewkowicz, 1988, 1996; Bahrick and Lickliter, 2000, 2004; Bahrick et al., 2002; Neil et al., 2006). There is also evidence for cross-modal facilitation, where stimuli in one modality increase the responsiveness to stimuli in other modalities (Lewkowicz and Lickliter, 1994; Lickliter et al., 1996; Morrongiello et al., 1998). On the other side, the ability to integrate unisensory signals in a statistically optimal fashion develops quite late, after 8–10 years of age (Gori et al., 2008, 2012; Nardini et al., 2008; Petrini et al., 2014; Dekker et al., 2015; Adams, 2016). Recent results show that during the first years of life, sensory modalities interact and communicate with each other and the absence of one sensory input impacts on the development of other modalities (Gori, 2015). According to the cross-sensory calibration theory, in children younger than 8–10 years old, the most robust sensory modality calibrates the other ones (Gori et al., 2008). This suggests that specific sensory modalities can be more suitable than others to convey specific information and hence to teach specific concepts. For example, it was observed that children use the tactile modality to perceive the size of objects, whereas the visual signal is used to perceive their orientation (Gori et al., 2008). It was also observed that when the motor information is not available, visual perception of size is impaired (Gori et al., 2012) and that when visual information is not available, tactile perception of orientation of objects is impaired (Gori et al., 2010). These results suggest that until 8–10 years of age, sensory modalities interact and shape each other. Then, multisensory technology that exploits multiple senses can be crucial to communicating specific concepts in a more effective way (i.e., having multiple signals available, the child can use the one which is most suitable for the task). Scientific evidence also suggests that it is not always true that the lack of one sensory modality is associated with an enhancement of the remaining senses (Rauschecker and Harris, 1983; Rauschecker and Knipert, 1994; Lessard et al., 1998; Röder et al., 1999, 2007; Voss et al., 2004; Lomber et al., 2010), but, in some cases, even the other not impaired senses are affected by the lack of the calibration modality (Gori et al., 2014; Finocchietti et al., 2015; Vercillo et al., 2015). For example, visually impaired children have impaired tactile perception of orientation (Gori et al., 2010); thus, touch cannot be used to communicate the orientation concept, and other signals, such as the auditory one, could be more suitable to convey this information.

We think that this scientific evidence should be reflected in teaching and learning practices, by introducing novel multisensory pedagogical methodologies grounded on it. In particular, we think that such scientific evidence supports an embodied and enactive pedagogical approach, using different sensory-motor signals and feedback (audio, haptic, and visual) to teach concepts to primary school children. For example, the use of sound associated with body movement could be an alternative way to teach visually impaired children the concept of orientation and angles. Such an approach would be more direct, i.e., natural and intuitive, since it is based on the

experience and on the perceptual responses to motor acts. Moreover, the use of movement for learning was shown to deepen and strengthen learning, retention, and engagement (Klemmer et al., 2006; Habib et al., 2016).

It should be noticed that sensory combination and sensory integration are implemented differently in the way multisensory signals are provided through technology. At the technological level, there is a difference between teaching a concept by using more than one modality (i.e., by adopting multiple alternative strategies and promoting multisensory combination) versus stimulating those modalities simultaneously by providing redundant sensory signals (thus promoting multisensory integration). In the technological area, Nigay and Coutaz (1993) classified multimodal interactive systems depending on their use of modalities and on whether modalities are combined (i.e., what in computer science is called multimodal fusion). In particular, they made a distinction between sequential, simultaneous, and composite multimodal interactive systems. The kind of multimodal interactive system, which is selected to provide multisensory feedback, depends on and affects the pedagogical paradigm and the way the learning process develops. More research is needed to get a deeper understanding of all the implications related to this choice, e.g., with respect to the concepts to teach, the needs of teachers and students, the optimal way of providing technological support, the learning outcomes, and so on.

## CHALLENGES

Multisensory technologies can help in overcoming the consolidated hegemony of vision in current educational practice. A too strong focus on one single sensory channel may compromise the effectiveness and personalization of the learning process. Moreover, a pedagogical approach based on one single modality may prevent the inclusion of children with impairments (e.g., with visual impairment).

More specifically, multisensory technologies can support the learning process by enhancing effectiveness, personalization, and inclusion. With respect to *effectiveness*, this may be affected by a wrong or an excessive usage of vision, which is not always the most suitable channel for communicating certain concepts to children.

As for *personalization*, a pedagogical methodology based almost exclusively on the visual modality would not consider the learning potential, and routes of access for learning in children, of exploiting the different modalities in ways that more comprehensively convey different kinds of information (e.g., the tactile modality is often better for perception of texture than vision). Moreover, we might speculate that a more flexible multisensory approach could highlight individual predispositions of children. It could be possible, for example, to observe a different individual tendency of preferring a specific learning approach for different children, demonstrating that specific sensory signals can be more useful for some children to learn specific concepts. For example, recent studies showed that musical training can be used as a therapeutic tool for



treating children with dyslexia (e.g., Habib et al., 2016). The temporal and rhythmic features of music could indeed exert a positive effect on the multiple dimensions of the “temporal deficit” that is characteristic of some types of dyslexia. Other specific examples are autism and Attention Deficit Hyperactivity Disorder (ADHD). There is solid evidence that multisensory stimuli improve the accuracy of decisions. Can multisensory technology or stimuli also improve attention or learning speed or retention? In our opinion, this issue is worthy of further investigation and found evidence in this direction would be crucial for designing effective technological support for children with developmental disorders (ADHD, autism, specific learning disabilities, dyslexia, and so on).

Concerning *inclusion*, the lack of vision in children with visual disability impacts, e.g., the learning of geometrical concepts that are usually communicated through visual representations and metaphors. A delay in the acquisition of cognitive skills in visually impaired children directly affects their social competence, producing in turn feelings of frustration that represent a risk for the development of personality and emotional competence (Thompson, 1941). The use of multisensory technology would allow having the same method for teaching to be used by sighted and blind children, thus naturally breaking barriers among peers and facilitating social interactions.

## AN OPPORTUNITY FOR MULTISENSORY TECHNOLOGIES

In our view, multisensory technologies are ideal for effectively supporting a pedagogical approach exploiting the best-suited sensory modality to teach a concept.

Multisensory technologies enable accurate and real-time mapping of motor behavior onto multiple facets of sound, music, tangible, and visual media, according to different strategies the teacher can select with great flexibility. Consider, for example, a recent technology-mediated learning activity we are developing for introducing geometric concepts, such as angles (Volta et al., 2018). In this activity, a child is asked to reproduce an angle by opening her arms. The child's arms represent the two sides of the angle and her head its vertex. Arms aperture is automatically measured by means of a Microsoft Kinect v.2 device, and the motor behavior is mapped onto multisensory feedback in real time. A visual or auditory feedback or both of them is provided. The visual feedback consists of the visual representation within a circle of the angle the child is doing (see **Figure 1**). Concerning the auditory feedback, while the child moves her arms, she can listen to a musical scale covering the full range of angle amplitude. If the child changes the aperture of her arms, the note in the scale – played by a string instrument – changes according to the movement. A long distance between the arms (i.e., a big angle) corresponds to a low-pitch note, whereas a short distance (i.e., a small angle) corresponds to a high-pitch note. Such a mapping is grounded on psychophysical evidence showing that a low pitch is associated with a big size and a high pitch is associated with a small size (Tonelli et al., 2017). If the child is able to

keep the same angle while rotating her arms, she listens to the same note with no changes in the auditory feedback, suggesting that angles are invariant under rotations. The teacher is provided with an interface enabling her to control the application (e.g., by selecting the angles that are proposed, the kind of feedback, several levels of difficulty, and so on). An initial and ongoing evaluation of this activity with children is suggesting that the proposed embodied representation of angles helps children in understanding angles and their properties (e.g., rotational invariance), even if more iterations of the development cycle are needed to address possible drawbacks (e.g., children get tired if asked to keep arms open for a too long time). While the angles activity implements a quite simple mapping of motor behavior onto visual and auditory feedback, more sophisticated approaches can be conceived. Multiple features of motor behavior can indeed be mapped onto multiple dimensions of sound morphology, including pitch, intensity, granularity, rhythm, and so on. While this is what usually happens when playing a musical instrument, technology makes it more flexible. Indeed, the teacher can choose which motor features are mapped onto which sound parameters and the child can quickly achieve a fine-grained control on the sound parameters, something that would require many years of practice with a traditional musical instrument. These issues have been debated for a long time in the literature of sound and music computing, see for instance, Hunt and Wanderley (2002), for a seminal work on this topic and the series of conferences on New Interfaces for Musical Expression<sup>2</sup>.

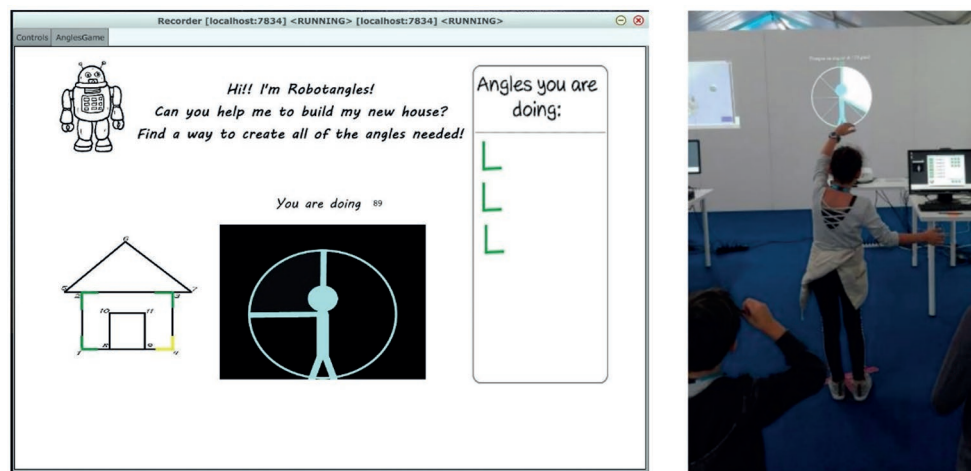
In our view, the adoption of an embodied and enactive pedagogical approach, tightly integrated with multisensory technology, would, therefore, foster effectiveness (for each specific concept, the most suited modality can be exploited) and personalization (flexibility for teachers and students) in the learning process. Moreover, inclusion can also take a great advantage: teaching can exploit the most suited substitutive modality for impaired children.

## GUIDELINES

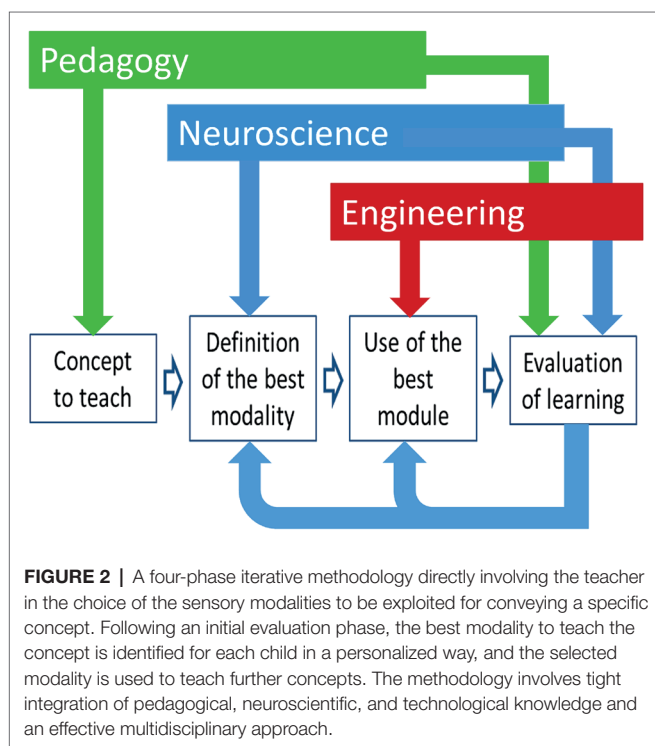
Since big opportunities most often entail likewise big risks, the introduction of multisensory technologies in the classroom needs to be careful. From our experience in technology-enhanced learning projects, we propose six golden rules we deem important for catching this opportunity and fully exploiting it.

1. *Ground technology on pedagogical needs.* Multisensory technologies should be tailored to the pedagogical needs of teachers. That is, they can help with teaching concepts that teachers specifically deem relevant in this respect. These could be concepts that are particularly difficult to understand for children or concepts that may enjoy communication through a sensory modality other than vision. We recently conducted a survey on over 200 math teachers. It was surprising for us to see that more than

<sup>2</sup><http://www.nime.org>



**FIGURE 1 |** The visual feedback provided by a technology-mediated learning activity featuring angles. A child makes an angle by opening her arms, the arms representing the sides of the angle, and the head its vertex. Arms aperture is measured by a Microsoft Kinect v.2 device and mapped in real time onto visual and/or auditory feedback. The activity was designed and developed by the Casa Paganini – InfoMus research centre at DIBRIS – University of Genoa and the U-Vip Unit at Istituto Italiano di Tecnologia in the framework of the EU-H2020-ICT weDRAW project.



**FIGURE 2 |** A four-phase iterative methodology directly involving the teacher in the choice of the sensory modalities to be exploited for conveying a specific concept. Following an initial evaluation phase, the best modality to teach the concept is identified for each child in a personalized way, and the selected modality is used to teach further concepts. The methodology involves tight integration of pedagogical, neuroscientific, and technological knowledge and an effective multidisciplinary approach.

75% of teachers agreed on the same concepts as the most difficult for children and the most appropriate for technological intervention.

2. *Ground technology on scientific evidence.* Multisensory technology should leverage on sensorial, perceptual, and cognitive capabilities children have according to scientific evidence. Concretely, for example, a technology able to detect specific motor behaviors in a target population of children (e.g., primary school) makes sense only if scientific
3. *Adopt an iterative design approach and rigorously assess learning outcomes.* Iterative design and refinement of technology are a critical component in the development cycle of interactive technologies. In case of multisensory technologies for education, this is crucial for successfully integrating technology in the classroom. Each iteration in the development process needs to rigorously assess learning outcomes (e.g., the speed of learning, longer-term outcomes like knowledge retention, student-centered outcomes like learner satisfaction, classroom behavior, and so on) and use this feedback to inform the next iteration.
4. *Make technology flexible and customizable.* This is a typical goal for technologies, but it assumes here a particular relevance. It means assessing (1) which is the preferred sensory modality for a child to learn a specific concept and (2) whether specific impairments require exploiting particular sensory modalities. As a side effect, technology may help with screening for behavioral problems and addressing them. For example, recent studies show that musical training can be used as a therapeutic tool for treating children with dyslexia (e.g., Curzon et al., 2009).
5. *Emphasize the role of the teacher.* In our view, technology does not replace the teacher. Rather, the teacher plays the central role of *mediator*. In an iterative methodology (see **Figure 2**), the teacher first chooses a concept to teach; then, following an initial evaluation phase, she identifies the best modality to teach it to each child and personalizes

technology to exploit the selected modality; she finally evaluates the outcomes of the learning process and adopts possible further actions. Moreover, design, development, and evaluation of technology should be obviously carried out in the framework of a participatory design process involving teachers and students (see Guideline 3).

6. *Promote cross-fertilization with the arts and human sciences.* Taking a rigorous scientific approach should not exclude the opportunity of getting inspiration from humanities, and in particular from arts. Recent initiatives (e.g., the EU STARTS platform<sup>3</sup>) witness the increased awareness of how art and science are two strongly coupled aspects of human creativity (Camurri and Volpe, 2016), as well as the impact of art on scientific and technological research. In case of multisensory technology for education, the extraordinary ability art has of conveying content by means of sound, music, and visual media provides, in our view, a significant added value.

## CONCLUSION

We developed and tested our approach in the framework of the weDRAW project<sup>4</sup>. This was an EU-H2020-ICT-funded project focusing on multisensory technologies for teaching math to primary school children. The final goal was to open a new teaching/learning channel based on multisensory interactive technology. The project represented an ideal testbed to assess the support of multisensory technology to learning math. More importantly, we think that the approach we outlined in this article can enable the development of a multisensory embodied

and enactive learning paradigm and of a teaching ecosystem that applies in the same way and provides the same opportunities to both typically developed and impaired children, thus breaking the barriers between them and fostering inclusion.

## ETHICS STATEMENT

This is a perspective paper, presenting our viewpoint on future research that will potentially involve human subjects. The perspective paper, however, is not based on any specific experimental study but is rather grounded on the existing literature and on our vision of future research.

## AUTHOR CONTRIBUTIONS

GV and MG equally contributed to both the concept of the article and its writing.

## FUNDING

This work was partially supported by the EU-H2020-ICT Project weDRAW. WeDRAW has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 732391.

## ACKNOWLEDGMENTS

We thank our colleagues participating in the weDRAW project for the insightful discussions.

<sup>3</sup><https://ec.europa.eu/digital-single-market/en/ict-art-starts-platform>

<sup>4</sup><http://www.wedraw.eu>

## REFERENCES

- Adams, W. J. (2016). The Development of Audio-Visual Integration for Temporal Judgements. *PLoS Comput. Biol.* 12:e1004865. doi: 10.1371/journal.pcbi.1004865
- Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029
- Bahrack, L. E., Flom, R., and Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Dev. Psychobiol.* 41, 352–363. doi: 10.1002/dev.10049
- Bahrack, L. E., and Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev. Psychol.* 36, 190–201. doi: 10.1037/0012-1649.36.2.190
- Bahrack, L. E., and Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: a developmental test of the intersensory redundancy hypothesis. *Cogn. Affect. Behav. Neurosci.* 4, 137–147. doi: 10.3758/CABN.4.2.137
- Baud-Bovy, G., and Balzarotti, N. (2017). "Using force-feedback devices in educational settings: a short review" in *Proceedings 1st ACM SIGCHI international workshop on multimodal interaction for education (MIE 2017)*, 14–21.
- Bruce, B. C., and Levin, J. A. (1997). Educational technology: media for inquiry, communication, construction, and expression. *J. Educ. Comput. Res.* 17, 79–102. doi: 10.2190/HPQ-4F3X-8M8Y-TVCA
- Camurri, A., and Volpe, G. (2016). The intersection of art and technology. *IEEE Multimedia* 23, 10–17. doi: 10.1109/MMUL.2016.13
- Clarke, J. J., and Yuille, A. L. (1990). *Data fusion for sensory information processing*. (Boston: Kluwer Academic).
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., and Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* 59, 661–686. doi: 10.1016/j.compedu.2012.03.004
- Curzon, P., McOwan, P. W., Cutts, Q. I., and Bell, T. (2009). Enthusing & inspiring with reusable kinaesthetic activities. *SIGCSE Bull.* 41, 94–98. doi: 10.1145/1595496.1562911
- Dekker, T. M., Ban, H., van der Velde, B., Sereno, M. I., Welchman, A. E., and Nardini, M. (2015). Late development of cue integration is linked to sensory fusion in cortex. *Curr. Biol.* 25, 2856–2861. doi: 10.1016/j.cub.2015.09.043
- Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a
- Ernst, M. O., and Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002
- Falloon, G. (2013). Young students using ipads: app design and content influences on their learning pathways. *Computers & Education* 68:505e521.
- Finocchietti, S., Cappagli, G., and Gori, M. (2015). Encoding audio motion: spatial impairment in early blind individuals. *Front. Psychol.* 6:1357. doi: 10.3389/fpsyg.2015.01357
- Ghahramani, Z., Wolpert, D. M., and Jordan, M. I. (1997). "Computational models of sensorimotor integration, in self-organization, computational maps and motor control" in *Self-organization, computational maps, and motor control*. eds. P. G. Morasso, and V. Sanguineti (Amsterdam: Elsevier), 117–147.

- Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisens. Res.* 28, 71–99. doi: 10.1163/22134808-00002478
- Gori, M., Del Viva, M., Sandini, G., and Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Curr. Biol.* 18, 694–698. doi: 10.1016/j.cub.2008.04.036
- Gori, M., Sandini, G., and Burr, D. (2012). Development of visuo-auditory integration in space and time. *Front. Integr. Neurosci.* 6:77. doi: 10.3389/fnint.2012.00077
- Gori, M., Sandini, G., Martinoli, C., and Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr. Biol.* 20, 223–225. doi: 10.1016/j.cub.2009.11.069
- Gori, M., Sandini, G., Martinoli, C., and Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain* 137, 288–293. doi: 10.1093/brain/awt311
- Gori, M., Tinelli, F., Sandini, G., Cioni, G., and Burr, D. (2012). Impaired visual size-discrimination in children with movement disorders. *Neuropsychologia* 50, 1838–1843. doi: 10.1016/j.neuropsychologia.2012.04.009
- Groff, J., and Mouza, C. (2008). A framework for addressing challenges to classroom technology use. *AACE J.* 16, 21–46.
- Gu, Y., Angelaki, D. E., and Deangelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* 11, 1201–1210. doi: 10.1038/nn.2191
- Habib, M., Lardy, C., Desiles, T., Commeiras, C., Chobert, J., and Besson, M. (2016). Music and dyslexia: a new musical training method to improve reading and related disorders. *Front. Psychol.* 7:26. doi: 10.3389/fpsyg.2016.00026
- Hicks, D., and Holden, C. (2007). *Teaching the global dimension: Key principles and effective practice*. (London: Routledge).
- Hunt, A., and Wanderley, M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound* 7, 97–108. doi: 10.1017/S1355771802002030
- Jiang, W., Jiang, H., Rowland, B. A., and Stein, B. E. (2007). Multisensory orientation behavior is disrupted by neonatal cortical ablation. *J. Neurophysiol.* 97, 557–562. doi: 10.1152/jn.00591.2006
- Jiang, W., Jiang, H., and Stein, B. E. (2002). Two corticotectal areas facilitate multisensory orientation behavior. *J. Cogn. Neurosci.* 14, 1240–1255. doi: 10.1162/089992902760807230
- Johnson, A. M., Jacovina, M. E., Russell, D. G., and Soto, C. M. (2016). “Challenges and solutions when using technologies in the classroom” in *Adaptive educational technologies for literacy instruction*. eds. S. A. Crossley and D. S. McNamara (NY, USA: Routledge, New York), 18.
- Joshi, R. M., Dahlgren, M., and Boulware-Gooden, R. (2002). Teaching reading in an inner city school through a multisensory teaching approach. *Ann. Dyslexia* 52, 229–242. doi: 10.1007/s11881-002-0014-9
- Katai, Z. (2011). Multi-sensory method for teaching-learning recursion. *Comput. Appl. Eng. Educ.* 19, 234–243. doi: 10.1002/cae.20305
- Katai, Z., and Toth, L. (2010). Technologically and artistically enhanced multi-sensory computer programming education. *Teach. Teach. Educ.* 26, 244–251. doi: 10.1016/j.tate.2009.04.012
- Klemmer, S. R., Hartmann, B., and Takayama, L. (2006). “How bodies matter: five themes for interaction design” in *Proceedings of the 6th conference on designing interactive systems*. (New York, NY, USA: ACM), 140–149.
- Kulik, J. A. (1983). Synthesis of research on computer-based instruction. *Educ. Leadersh.* 41, 19–21.
- Landy, M. S., Banks, M. S., and Knill, D. C. (2011). “Ideal-observer models of cue integration” in *Sensory cue integration*. eds. J. Trommershauser, K. Kording, and M. S. Landy (Oxford: Oxford University Press), 5–29.
- Lessard, N., Paré, M., Lepore, F., and Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395, 278–280. doi: 10.1038/26228
- Lewkowicz, D. J. (1988). Sensory dominance in infants: 1. Six-month-old infants’ response to auditory-visual compounds. *Dev. Psychol.* 24, 155–171. doi: 10.1037/0012-1649.24.2.155
- Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1094–1106. doi: 10.1037/0096-1523.22.5.1094
- Lewkowicz, D. J., and Lickliter, R. (eds.) (1994). *The development of intersensory perception: Comparative perspectives*. (Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc.).
- Lickliter, R., Lewkowicz, D. J., and Columbus, R. F. (1996). Intersensory experience and early perceptual development: the role of spatial contiguity in bobwhite quail chicks’ responsiveness to multimodal maternal cues. *Dev. Psychobiol.* 29, 403–416. doi: 10.1002/(SICI)1098-2302(199607)29:5<403::AID-DEV1>3.0.CO;2-S
- Lomber, S. G., Meredith, M. A., and Kral, A. (2010). Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nat. Neurosci.* 13, 1421–1427. doi: 10.1038/nn.2653
- Moreno, R., and Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *J. Educ. Psychol.* 91, 358–368. doi: 10.1037/0022-0663.91.2.358
- Morronegello, B. A., Fenwick, K. D., and Chance, G. (1998). Cross-modal learning in newborn infants: inferences about properties of auditoryvisual events. *Infant Behav. Dev.* 21, 543–554. doi: 10.1016/S0163-6383(98)90028-5
- Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Curr. Biol.* 18, 689–693. doi: 10.1016/j.cub.2008.04.021
- Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., and Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Dev. Sci.* 9, 454–464. doi: 10.1111/j.1467-7687.2006.00512.x
- Nigay, L., and Coutaz, J. (1993). “A design space for multimodal systems: concurrent processing and data fusion” in *Proceedings of the INTERACT’93 and CHI’93 conference on human factors in computing systems (CHI’93)*. (New York, NY, USA: ACM), 172–178.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. (New York, USA: Basic Books Inc.).
- Petrini, K., Remark, A., Smith, L., and Nardini, M. (2014). When vision is not an option: children’s integration of auditory and haptic information is suboptimal. *Dev. Sci.* 17, 376–387. doi: 10.1111/desc.12127
- Philip, T. M. (2017). Learning with mobile technologies. *Commun. ACM* 60, 34–36. doi: 10.1145/2976735
- Raposo, D., Kaufman, M. T., and Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* 17, 1784–1792. doi: 10.1038/nn.3865
- Raposo, D., Sheppard, J. P., Schrater, P. R., and Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *J. Neurosci.* 32, 3726–3735. doi: 10.1523/JNEUROSCI.4998-11.2012
- Raskind, M., Smedley, T. M., and Higgins, K. (2005). Virtual technology: bringing the world into the special education classroom. *Interv. Sch. Clin.* 41, 114–119. doi: 10.1177/10534512050410020201
- Rauschecker, J. P., and Harris, L. R. (1983). Auditory compensation of the effects of visual deprivation in the cat’s superior colliculus. *Exp. Brain Res.* 50, 69–83. doi: 10.1007/BF00238233
- Rauschecker, J. P., and Kniepert, U. (1994). Auditory localization behaviour in visually deprived cats. *Eur. J. Neurosci.* 6, 149–160. doi: 10.1111/j.1460-9568.1994.tb00256.x
- Rizzo, A., Raheb, K. E., Whatley, S., Velaetis, M., Zanon, M., Camurri, A., et al. (2018). “WhoLoDancE: whole-body interaction learning for dance education” in *Proceedings of the workshop on cultural informatics at the EUROMED international conference on digital heritage 2018 (EUROMED2018)*, 41–50.
- Röder, B., Kusmierek, A., Spence, C., and Schicke, T. (2007). Developmental vision determines the reference frame for the multisensory control of action. *Proc. Natl. Acad. Sci. USA* 104, 4753–4758. doi: 10.1073/pnas.0607158104
- Röder, B., Teder-Sälejärvi, W., Sterr, A., Röslér, F., Hillyard, S. A., and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature* 400, 162–166. doi: 10.1038/22106
- Rowland, B. A., Jiang, W., and Stein, B. E. (2014). Brief cortical deactivation early in life has long-lasting effects on multisensory behavior. *J. Neurosci.* 34, 7198–7202. doi: 10.1523/JNEUROSCI.3782-13.2014
- Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., and Kato, H. (2014). Augmented reality learning experiences: survey of prototype design and evaluation. *IEEE Trans. Learn. Technol.* 7, 38–56. doi: 10.1109/TLT.2013.37
- Sheppard, J. P., Raposo, D., and Churchland, A. K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *J. Vis.* 13, pii: 4. doi: 10.1167/13.6.4
- Stein, B. E., and Meredith, M. A. (1993). *The merging of the senses*. (Cambridge, Mass: MIT Press).



- Stein, B. E., Stanford, T. R., and Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nat. Rev. Neurosci.* 15, 520–535. doi: 10.1038/nrn3742
- Supalo, C. A., Wohlers, H. D., and Humphrey, J. R. (2011). Students with blindness explore chemistry at “camp can do”. *J. Sci. Educ. Stud. Disabil.* 15, 1–9. doi: 10.14448/jesd.04.0001
- Thompson, J. (1941). Development of facial expression of emotion in blind and seeing children. *Arch. Psychol.* 37, 1–47. doi: 10.1038/nrn3742
- Tonelli, A., Cuturi, L. E., and Gori, M. (2017). The influence of auditory information on visual size adaptation. *Front. Neurosci.* 11:594. doi: 10.3389/fnins.2017.00594
- Varni, G., Volpe, G., Sagoleo, R., Mancini, M., and Lepri, G. (2013). “Interactive reflexive and embodied exploration of sound qualities with BeSound” in *Proceedings of the 12th international conference on interaction design and children (IDC2013)*, 531–534.
- Vercillo, T., Milne, J. L., Gori, M., and Goodale, M. A. (2015). Enhanced auditory spatial localization in blind echolocators. *Neuropsychologia* 67, 35–40. doi: 10.1016/j.neuropsychologia.2014.12.001
- Volta, E., Alborno, P., Gori, M., and Volpe, G. (2018). “Designing a multisensory social serious-game for primary school mathematics learning” in *Proceedings 2018 IEEE games, entertainment, media conference (GEM 2018)*, 407–410.
- Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J. P., and Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr. Biol.* 14, 1734–1738.
- Walling, D. R. (2014). *Designing learning for tablet classrooms: Innovations in instruction—using tablet technology for multisensory learning*. (Basel, Switzerland: Springer International Publishing).
- 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 Volpe and Gori. 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.



# Building Blocks of Mathematical Learning: Virtual and Tangible Manipulatives Lead to Different Strategies in Number Composition

Ana Cristina Pires<sup>1,2\*</sup>, Fernando González Perilli<sup>2,3</sup>, Ewelina Bakala<sup>4</sup>, Bruno Fleisher<sup>2,3</sup>, Gustavo Sansone<sup>5</sup> and Sebastián Marichal<sup>6</sup>

<sup>1</sup> LASIGE, Faculty of Sciences, Universidade de Lisboa, Lisbon, Portugal, <sup>2</sup> Faculty of Psychology, Center for Basic Research in Psychology, Universidad de la República, Montevideo, Uruguay, <sup>3</sup> Faculty of Information and Communication, Universidad de la República, Montevideo, Uruguay, <sup>4</sup> Faculty of Engineering, Computer Science Institute, Universidad de la República, Montevideo, Uruguay, <sup>5</sup> Faculty of Architecture, Universidad de la República, Montevideo, Uruguay, <sup>6</sup> Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Firat Soylu,  
University of Alabama, United States

### Reviewed by:

Kasia Muldner,  
Carleton University, Canada  
Jennifer M. Zosh,  
Pennsylvania State University,  
United States

### \*Correspondence:

Ana Cristina Pires  
acdpires@di.fc.ul.pt

### Specialty section:

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

**Received:** 16 August 2018

**Accepted:** 23 July 2019

**Published:** 06 September 2019

### Citation:

Pires AC, González Perilli F, Bakala E, Fleisher B, Sansone G and Marichal S (2019) Building Blocks of Mathematical Learning: Virtual and Tangible Manipulatives Lead to Different Strategies in Number Composition. *Front. Educ.* 4:81. doi: 10.3389/feduc.2019.00081

Multiple kinds of manipulatives, such as traditional, virtual, or technology-enhanced tangible objects, can be used in primary education to support the acquisition of mathematical concepts. They enable playful experiences and help children understand abstract concepts, but their connection with cognitive development is not totally clear. It is also not clear how virtual and physical materials influence the development of different strategies for solving instructional tasks. To shed light on these issues, we conducted a 13-day intervention with 64 children from first grade, divided into three groups: Virtual Interaction (VI), Tangible Interaction (TI), and Control Group (CO). The VI group played a fully digital version of a mathematics video game and the manipulation of the blocks took place on the tablet screen. The TI group played the same video game with digitally augmented tangible manipulatives. Finally, the CO group continued with their classroom curricular activities while we conducted the training, and only participated in the Pre and Post-Test evaluations. Our results highlighted that the use of tangible manipulatives led to a positive impact in children's mathematical abilities. Of most interest, we recorded children's actions during all the training activities, which allowed us to achieve a refined analysis of participants' operations while solving a number composition task. We explored the differences between the use of virtual and tangible manipulatives and the strategies employed. We observed that the TI group opted for a greater number of blocks in the number composition task, whereas the VI group favored solutions requiring fewer blocks. Interestingly, those children whose improvement in mathematics were greater were the ones employing a greater number of blocks. Our results suggest that tangible interactive material increases action possibilities and may also contribute to a deeper understanding of core mathematical concepts.

**Keywords:** digital manipulatives, tangible manipulatives, technology-enhanced learning activities, mathematics, additive composition

## 1. INTRODUCTION

Learning mathematics at an early age is fundamental to ensuring academic success in STEM (science, technology, engineering, and mathematics) disciplines and maximizing future integration into professional life (Wang and Goldschmidt, 2003). Research has been concerned with how to foster this core cognitive ability and enable a deep understanding of mathematical concepts. This research explores how virtual and tangible manipulatives can be used to strengthen math learning at 6 years of age.

In the current study, we used the activity of composing and decomposing sets of manipulatives representing numbers, an exercise that has been traditionally practiced with concrete material in order to foster an understanding of numerosity (Geary et al., 1992; Morin and Franks, 2009). We focused on a set of three properties (additive composition, commutativity, and associativity) and the mastery of the basic number combinations. Additive composition is the knowledge that larger sets are made up of smaller sets; the commutative property implies that changing the order of the operands doesn't affect the result; the associative property allows us to add (or multiply) numbers, no matter how the factors are grouped  $[(a + b) + c = a + (b + c)]$ ; while mastering the basic number combinations leads to understanding how numbers can be composed. These properties are crucial for cardinality and number concept acquisition; and lead to the development of key strategies in arithmetical problem solving, such as addition and subtraction (Fuson, 1992; Verschaffel et al., 2007).

In mathematics curricula, teaching is frequently supported by tangible objects (three-dimensional models of geometrical shapes, etc.) that help young students to better understand abstract concepts, for instance in the acquisition of cardinality (Geary et al., 1992; Morin and Franks, 2009). The pioneer in this tradition was Maria Montessori who developed materials for geometry and mathematics specifically aimed at providing children with autonomy during the learning process (Montessori, 1917). Georges Cuisenaire, in turn, created a special set of tiles for arithmetics learning known as Cuisenaire rods (Cuisenaire, 1968). His proposal was based on the relationship between size and number and exploited the possibility of different spatial arrangements to exemplify mathematical principles like number composition. A new version of these materials can be found in Singapore Math's tiles (Wong, 2009; Wong and Lee, 2009); which is considered one of the more influential methods for teaching basic mathematics nowadays (Deng et al., 2013).

Following this vein, the acquisition of the number concept—one of the building blocks of mathematical learning—would benefit from direct interaction with objects (Dienes, 1961; Chao et al., 2000; Anstrom, 2006; McGuire et al., 2012). Interaction with objects may facilitate the passage from a concrete construal (I can see/manipulate three things in front of me) toward an abstract one ( $3 = ***$ ). This transformation begins with a process which is strongly based on perceptual, non verbal operations and turns into a symbolic one supported by an abstract association (Feigenson et al., 2004). The first stage has to do with the understanding that a given group of objects has a certain quantity

of components (Gelman and Gallistel, 1978); the second with associating this quantity (of objects) to an exact number and its symbolic expression, and then understanding that any time the number is seen or heard it means that an exact quantity is being referred to (Kilpatrick et al., 2001).

The sensitivity to numerosity is improved gradually as the infant develops (Izard et al., 2009). Infants even just a few hours old are already sensitive to numerosity (e.g., Antell and Keating, 1983; Izard et al., 2009). Allegedly, this is possible due to two innate parallel number systems (see Feigenson et al., 2004; for a review see Piazza, 2010): an object file system (Feigenson and Carey, 2003) which accounts for the immediate identification of a discrete quantity of elements—subitizing (Kaufman and Lord, 1949)—and is limited by the capability to attend to different objects at the same time; and an approximate number system (ANS) which accounts for a non-symbolic continuous numerical representation involving large numbers (Gallistel and Gelman, 1992; Dehaene, 2011).

Nevertheless, children are not able to explicitly identify simple quantities involving numbers from 1 to 4 until 4 years old, and up to 5 until 5 years old. To do so, different skills must be developed such as counting and conceptual subitizing; the combination of two “subitizable” numbers, for e.g., recognizing the presence of a 3 (\*\*\*) and a 4 (\*\*\*\*) and implicitly composing a set of 7 (\*\*\*\*\*) (Steffe and Cobb, 1988; Clements, 1999). Toddlers recognize that sets can be combined in different ways, but this understanding is based on nonverbal, perceptual processes (Sophian and McCorgay, 1994; Canobi et al., 2002). Commutativity is only acquired later between 4 and 5 years old, as also the understanding that commutativity of added groups leads to associativity (Gelman and Gallistel, 1978; Canobi et al., 2002). Thus, associativity reflects conceptual reasoning about how groups can be decomposed and recombined (Sarama and Clements, 2009). Further, as children learn basic number combinations, they can master a broad set of heuristics when faced with addition and subtraction problems.

To foster the conceptualization of unit items children may rely on hand actions such as pointing or grasping (Steffe and Cobb, 1988). For instance, in the case of subtraction, small children often represent the minuend with the fingers (or objects) and fold their fingers (or remove objects) for the value of the subtrahend (Groen and Resnick, 1977; Siegler, 1984). In fact, most children cannot solve complex numerical problems without the support of concrete objects until 5.5 years old (Levine et al., 1992). Later on, children acquire retrieval strategies, accessing results directly from long term memory (Rathmell, 1978; Steinberg, 1985; Kilpatrick et al., 2001). For this to be possible, children need to master basic number combinations (Baroody and Tiilikainen, 2003), but also understand associativity (Sarama and Clements, 2009). Children typically progress throughout three phases to achieve mastery on basic number combinations: (a) Counting strategies—using object counting (e.g., with blocks, fingers) or verbal counting (b) Reasoning strategies—using known information (facts and relationships) to deduce the answer of an unknown combination; (c) Mastery-efficient responses [i.e., fast and accurate (Kilpatrick et al., 2001)].

Children's addition and subtraction strategies also evolve during childhood. For instance, in order to solve  $9 + 8$ , 4 to 5-year-old children would count from 1 to 9 for the first addend and then from 9 to 17 for the total sum ("counting all strategy"; Fuson, 1992; Verschaffel et al., 2007). Later on between 5 and 6 years old children would develop the more refined strategy of "counting on" in which the count starts from the cardinal of the larger addend (i.e., from 9 to 17; Carpenter and Moser, 1982; Siegler and Jenkins, 2014). More sophisticated part-whole strategies are developed with the achievement of associativity and the knowledge of how numbers from 1 to 10 can be composed (6–7 years old; Canobi et al., 2002). To solve  $9 + 8$  children would be able to retrieve that  $9 + 1$  is one of the forms to compose 10, and then solve the problem by the easier  $10 + 7$  (also retrieving that  $8 - 1$  equals 7; Carpenter and Moser, 1984; Fuson, 1992; Miura and Okamoto, 2003).

Interaction with objects may supports the development of different strategies by diminishing cognitive load and freeing up working memory, given that the perceived entities are cognitively available through the objects that represent them in space (Manches and O'Malley, 2016). Object manipulation gives rise to operations that can work as analogies of abstract operations. For example, joining 2 elements to a group of another 3 forms a new group of 5. This concrete activity would be a metaphor of act of addition:  $2 + 3 = 5$ . These conceptual metaphors work as scaffolding that allows children to grasp abstract ideas such as commutativity or associativity (Manches and O'Malley, 2016).

With the appearance of digital technologies, researchers have been exploring how the manipulation of digital (Yerushalmy, 2005; Moyer-Packenham and Westenskow, 2013) and/or technology-enhanced concrete material (Tangible User Interfaces or TUIs; Manches, 2011) can benefit learning processes, finding promising results (see Sarama and Clements, 2016). Beyond the encouraging results obtained in several technology-based interventions, it has been claimed that the application of digital technology in the classroom posits the risk of replacing rich physical interactions with the environment by much more constrained interactions such as the use of the mouse-keyboard or multi-tactile interfaces (Bennett et al., 2008). In this vein, theories like constructivism, embodied cognition (Wilson, 2002; Anderson, 2003) and physically distributed learning (Martin and Schwartz, 2005) support the idea that physical interaction plays a key role in the learning process (Antle and Wise, 2013; for a review in this matter see Sarama and Clements, 2016).

In this study, we focus on the kinds of actions virtual and physical manipulatives offer and their impact on numerical learning. On one hand, interaction with virtual manipulatives is limited to dragging objects on the screen, but it still allows children to displace, join and isolate objects as traditional manipulatives allow (Moyer-Packenham and Westenskow, 2013). On the other hand, classic manipulatives offer interactive advantages (to grasp the object, for instance) that could have relevant consequences for educational activity (Martin and Schwartz, 2005; Manches and O'Malley, 2016). Several

studies have been dedicated to this comparison, providing results which are slightly favorable to physical manipulatives (Martin and Schwartz, 2005; Schwartz et al., 2005; Klahr et al., 2008).

Technology-enhanced tangible manipulatives offer several advantages when compared with traditional or virtual manipulatives (Moyer-Packenham and Westenskow, 2013). They allow autonomous and active learning by using physical material and enable us to record a child's performance. In addition, they enable us to explore which kind of actions are relevant in specific learning activities. Importantly for the present research, our system permits analyzing and comparing the use of physical and virtual manipulatives to solve a task of additive composition. This comparison is of special theoretical interest given that it makes possible to explore the role of physicality/three-dimensionality in learning mathematics. In other words, the present research aims to investigate if it is indispensable that objects may be grasped, lifted, and explored or would it be enough to interact with virtual manipulatives? And specifically, we ask how the objects' affordances (i.e., the possibility to grasp physical objects or drag virtual ones) will shape and constrain children's composing strategies.

## 2. MATERIALS AND METHODS

### 2.1. Participants

We recruited participants from one state school in Montevideo (Uruguay) with a medium-high sociocultural status consisting of 64 children (three classrooms) from first grade. All children had an informed consent form signed by their parents or legal guardians. A research protocol was approved by the Local Research Ethics Committee of the Faculty of Psychology, and is in accordance with the 2008 Helsinki Declaration. We employed a quasi-experimental design and each classroom became one of the following experimental groups: Control (CO), Virtual Interaction (VI), and Tangible Interaction (TI).

Four children (two from the VI group and another two from the TI group) failed to correctly answer 25% of the trials in our training game. Therefore, we performed subsequent analyses with the remaining 60 children (33 girls and 27 boys). Group descriptive information is shown in **Table 1**. We examined the effect of age and sex by conducting separated *t*-tests on assessment scores, but we did not find any effect.

**TABLE 1 |** Mean and standard deviations at pre- and post-tests by groups.

	<i>n</i>	Age (years)	Sex (*girls)	TEMA-3	
				Pre	Post
Passive Group (PA)	20	6.6 (0.3)	13	25.6 (5.7)	28.8 (4.6)
Virtual Interaction Group (VI)	20	6.8 (0.5)	11	31.8 (9.6)	35.1 (9.3)
Tangible Interaction Group (TI)	20	6.8 (0.6)	11	30.2 (10.3)	34.4 (10.5)



## 2.2. Procedure

To evaluate the impact of both game modalities in the acquisition of mathematical abilities, we planned an intervention with three phases. A first and last phase of evaluations (Pre- and Post-Test), and a training of 13 days in between.

### 2.2.1. Pre-test

To evaluate children's mathematical abilities before and after training we used the third edition of the standardized Test of Early Mathematics Ability (TEMA-3, Bliss, 2006) for children between 3 and 8 years of age. The test was verbally administered and consisted of 72 items to assess: counting ability, number comparison facility, numeral literacy, mastery of number facts, basic calculation skills, and understanding of mathematical concepts. This test has high content validity (Baroody, 2003) and high reliability ranging from 0.82 to 0.97. Indeed, we found a high test-retest reliability measured by calculating TEMA-3 correlation between Pre-Test and Post-Test measures across children within each training group (TI: 0.94; VI: 0.94; CO: 0.78). We calculated scores by the sum of all the correct answers (taking into account ceiling and floor effects that are part of the test administration). Two trained evaluators conducted the evaluation and it took about 30 min per participant. This phase took one week, with 12 children evaluated per day.

### 2.2.2. Training/Playing

The three classes selected to participate in the study continued with their regular formal learning activities as part of the school curriculum. Apart from the fact that each class had a different teacher, teachers followed the same program and protocol, and were committed to giving the same math curricula information for the three classes. Both the TI and VI group played over 13 days (3 weeks). Sessions had a duration of 20 min each, from Monday to Friday. Two researchers were present in every session to help with any technical problems that may have arisen. In the first session, we introduced the game dynamics and made explicit the relation between size and value of each tangible and virtual block to facilitate effective use of manipulatives. The CO group continued with their regular curricular activities while the other two groups had 20 min per day of training. The CO group only participated in the Pre- and Post-Tests assessments.

### 2.2.3. Post-test





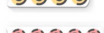
The same evaluators assessed the groups again with TEMA-3 and the scores were analyzed in the same manner as in the Pre-Test evaluation.

## 2.3. Training Game BrUNO

The video game BrUNO was developed to give the learning activity a more attractive and playful format. We took gamification theory into consideration in order to incorporate some gamification elements in BrUNO, such as: microworlds, a main-character, a tutorial, several types of prizes, and funny sounds. During the development of BrUNO, we carried out two informal user tests to inform the game design (Marichal et al., 2017a).

BrUNO is a video game designed to work on additive composition. Children played BrUNO by using five types of blocks whose length and color were associated with their value (see **Figure 1**). The block of 1 represents the number "1"; the block of 2 represents the number "2," and so forth until 5. Each block has a different length which is proportional to the value that it represents).

To facilitate visual recognition of the location of the number required to build, a horizontal or vertical number line (depending on the scenario) is shown on the screen (see **Figure 2**). It is known that as numerosity develops, a hierarchical mental representation of how numbers should be ordered arises in the form of a number line. This line, which is based on a spatial analogy, represents the numbers from lowest to highest and locates them according to their cardinality. Thus, to reinforce this mental representation and to facilitate the additive composition task, we presented a number line to guide the players while they compose the required number. It helps to count the missing/spare units and deduce

Block	Block value	Digital block dimensions	Tangible block dimensions	Color
	1	40px x 40px	16mm x 16mm	yellow
	2	80px x 40px	32mm x 16mm	green
	3	120px x 40px	48mm x 16mm	blue
	4	160px x 40px	64mm x 16mm	orange
	5	200px x 40px	80mm x 16mm	red

**FIGURE 1** | Block values, dimensions, and color.



**FIGURE 2** | Fully virtual version of BrUNO. Prize placed in number three (as indicated by the orange color). The player has already introduced 1 block of value 2. To reach the prize, he must add one block of value 1. In this example, a horizontal number line is present to help children locating numbers and to help in adding and subtracting operations.

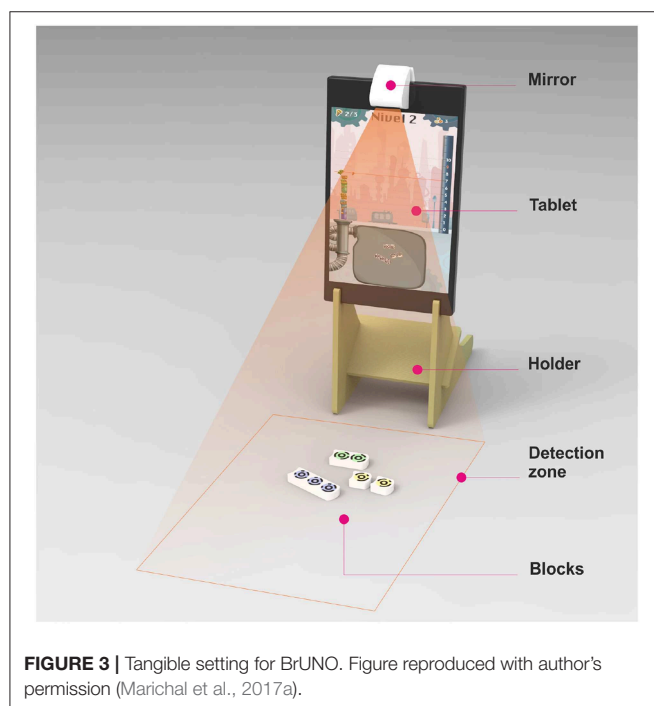
how the target number can be correctly composed. If the child has to build the number 4 and she has already put one block of 3, she can observe that the game character is 1 unit away from the prize and compose the target number by adding the block of 1. This way, the child can learn that  $3 + 1 = 4$ . Additionally, the game helps to demonstrate that, for example, the distance between 1 and 3 is the same as between 21 and 23—a fact that is not so obvious for young children (Siegler and Booth, 2004).

We developed two conditions for the evaluation of manipulatives: the Tangible Interaction Group (TI) and the Virtual Interaction Group (VI). In both cases, children played BrUNO, but the interaction with the blocks differed. In the first case, children manipulated technology-enhanced tangible blocks, and in the second case, virtual blocks.

### 2.3.1. Tangible Interaction Device

We designed a low cost tangible interaction device named CETA (Marichal et al., 2017a), with three main components (see **Figure 3**): a mirror that changes the webcam's viewing direction, allowing the system to detect objects over the table; a wooden holder that keeps the tablet vertically in portrait orientation; and a set of tangible blocks of different sizes similar to Cuisenaire Rods (representing numbers from 1 to 5; see **Figure 1**).

We used the webcam of the tablet and a mirror to capture the image of the surface in front of the tablet holder in real-time. This image is constantly analyzed to detect blocks in the detection zone (for more details see Marichal et al., 2017b). The limits of the detection zone are determined by the webcam hardware and height of the holder. Blocks outside the detection zone are not visible to the computer vision system.



**FIGURE 3 |** Tangible setting for BrUNO. Figure reproduced with author's permission (Marichal et al., 2017a).

We designed a set of 25 blocks for 3D printing. The handling capabilities of the children at target age, the dimensions of the detection zone of the computer vision system, and the numeric quantities required by the different game challenges determined the dimensions of the blocks. All blocks contain magnets at their extremities, providing an affordance that increases the probability of joining blocks imitating the number line representation. Every block has a positive and a negative extremity. The concave and convex block's terminations constrain the way it can be joined. On the top face of each block we placed a set of colored markers (TopCodes; Horn, 2012) used by the computer vision system. The number of markers on each block corresponds to the block value.

### 2.3.2. Virtual Interaction Device

The virtual version allows to play BrUNO without CETA device. The blocks are virtual and the child has to place them in the detection zone to submit its answer to the system (**Figure 2**).

### 2.3.3. Data Collection

We recorded the children's actions to trace the quantity and the type of blocks employed in children's solutions over time. This allowed us to analyze the game strategies developed by each group and follow the performance of every single participant. After each response our system recorded the following data: (1) the number required to form, (2) the number actually formed, and (3) the blocks used to form the number.

We assumed that if the child wanted to respond with two blocks but put the first block in the detection zone while looking for the other, then we should develop a strategy to avoid considering this incomplete answer as a child's final solution. Thus, to avoid recording partial solutions we implemented what we call "action submit," which consists of two steps. The first step is to wait for a stable solution. By stable solutions, we mean invariant responses by children for 1.5 s meaning that the blocks placed in the detection zone were not moved for 1.5 s and no blocks were added or removed. If this condition was completed, then we move to the second step in which the game character prepares itself for 1 s to execute the movement. If, during this time the child changed his or her answer, the time counter resets and "action submit" starts over again. If the answer did not change, the game character moves and the system records the blocks that composed the child's solution. To avoid duplicate responses (e.g., the child leaves the blocks in the detection zone and goes to the bathroom) we only registered the solutions that differed from the last recorded solution.

## 3. RESULTS

### 3.1. Differences Between Groups

To test the effect of playing our training game over 13 sessions, we assessed the children's mathematics performance using TEMA-3 before and after training or without training as in the case of the CO group.

While we had a quasi-experimental design in which the groups were non-randomized at baseline, there were no significant differences between groups on Pre-Test,  $p = 0.84$ . To test for

conditional differences, we used an ANCOVA with the Post-Test scores as the dependent variable, the Pre-Test as the covariate, and the Group as the independent variable. ANCOVA is advocated in this type of context because it controls for minor variations in the Pre-Test scores (Oakes and Feldman, 2001; Schneider et al., 2015). The assumptions of the ANCOVA were satisfied (as noted above, the covariate levels did not differ between conditions, and homogeneity of slopes held, as verified by running an ANOVA and customizing the model to include the interaction between the covariate and independent variable,  $p = 0.5$ ). The ANCOVA identified a significant effect of Group,  $F_{(2,54)} = 20.9$ ,  $p < 0.001$ ,  $r = 0.44$ . We followed up this analysis with pairwise comparisons between Post-Test scores adjusted by the ANCOVA with the baseline Pre-Test scores. Both experimental groups obtained higher Post-Test scores than the control group ( $VI_{Mean}: 32.54$ ,  $VI_{SD} = 0.77$ ;  $TI_{Mean}: 33.27$ ,  $TI_{SD} = 0.74$  and  $CO_{Mean}: 30.93$ ,  $CO_{SD} = 0.86$ ). However, only Post-Tests scores significantly differed when comparing TI vs CO ( $p = 0.044$ ). We found no other significant effects between groups.

### 3.2. Virtual and Tangible Interaction Groups and the Minimum Blocks Coefficient (MBC)

We focused on the possible problem-solving strategies employed by the children when resolving the number composition task, and how the type of interaction could have affected their actions. To do so, we carried out exploratory analysis using participants' log files. It allowed us to observe which blocks were used to compose each number by all the participants, at every successful trial.

Firstly, we analyzed whether the number of blocks used to build the correct solution was different across groups. For example, to build the number 3, it is possible to use three blocks of 1 ("1-1-1"), one block of 1 and one block of 2 ("1-2"), or directly use one block of 3 ("3"). To evaluate how close the child was to using the minimum number of blocks that were necessary to build a number (one block in the case of numbers from 1 to 5, two blocks in case of numbers from 6 to 10, or three blocks if the number is greater than 10), we developed a score called the "Minimum Blocks Coefficient" (MBC). MBC is a metric that allows us to observe the different solutions in composing numbers while training additive composition. We aim to explore how children compose numbers using different types of manipulatives. For each correct solution it takes the minimum number of blocks necessary to build the number requested, and divides it by the number of blocks actually used. For example, in the case of number 3 the variant "1-1-1" becomes the score  $1/3 = 0.33$ , because just one block is necessary to build the number (block of 3), and in reality, three blocks were used. The combination "1-2," becomes  $1/2 = 0.5$ , and "3," becomes the score of 1.0. To calculate the MBC for one particular number and one particular group (TI or VI), we take all the correct solutions of the number formed by the participants of the group and calculate the mean value. Error rates were not analyzed because we observed that the tangible system required more time for the physical manipulation and during that time some partial

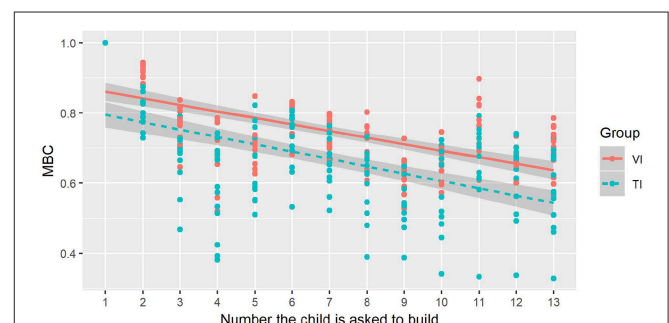
solutions were recorded as errors before the child's final answer. For example, if the child wanted to respond with two blocks, but he or she put the first block in the detection zone while looking for the other and no changes occur in the detection zone for 2.5 s, the system registered the child's uncompleted solution as a response (error in this case). The algorithm is explained with more detail in the section "2.3.3." For the aforementioned reasons we decided to only analyze the correct answers, so we were confident that we analyzed explicitly correct answers rather than random solutions.

#### 3.2.1. Minimum Blocks Coefficient by Numbers (1–13)

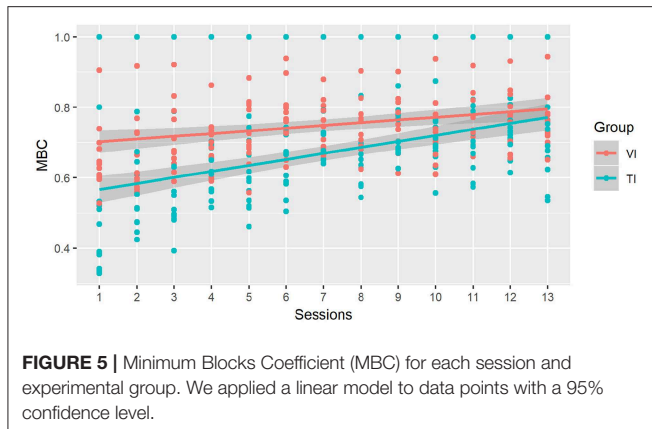
We applied a two-way ANOVA considering the MBC as the dependent variable and Group and Numbers as the independent variables. Numbers is the variable that represents the number the child is asked to build. We divided all the Numbers that appear in the game (1–13) into three ranges based on the theoretical MBC that could be used for those numbers. Specifically, the theoretical MBC for numbers ranged from 1 to 5 is one block (i.e., they have the possibility to respond with a minimum of one block); for the numbers ranged 6–10 is two (i.e., they have the possibility to respond with a minimum of two blocks) and for the numbers ranged from 11 to 13 is three blocks (i.e., they have the possibility to respond with a minimum of three blocks).

The results showed that the type of manipulatives (TI or VI group) [ $F_{(1,126)} = 6.21$ ,  $p = 0.014$ ,  $r = 0.076$ ] and the Number [ $F_{(2,126)} = 10.8$ ,  $p < 0.001$ ,  $r = 0.060$ ] (see **Figure 4**) significantly influenced the MBC. We found no further interaction. The TI group used significantly more pieces (lower MBC) comparing with the VI group ( $TI_{Mean} = 0.65$ ,  $TI_{SD} = 0.19$ ,  $VI_{Mean} = 0.72$ ,  $VI_{SD} = 0.15$ ). These differences between TI and VI may be a result of the diverse composing strategies used when solving the number composition task.

Considering the variable Number, the number of blocks used were significantly fewer for the numbers ranging from 1 to 5 compared to the numbers ranging from 6 to 10 ( $p = 0.0002$ ) and also compared to the numbers ranging from 11 to 13 ( $p = 0.0003$ ).



**FIGURE 4 |** The Minimum Blocks Coefficient (MBC) for each number the child was asked to build. We applied a linear model to data points with a 95% confidence level for each Experimental Group: Virtual Interaction (VI) and Tangible Interaction (TI).



### 3.2.2. Minimum Blocks Coefficient Over Time

Participants reduced the number of blocks used during the 13 sessions that our intervention lasted (see **Figure 5**). We found a significant positive correlation ( $ps < 0.0001$ ) between the MBC and sessions for VI (0.84) and for TI (0.87) groups. We also explored whether the number of blocks employed was significantly different at different moments of our intervention by analysing the MBC Mean for the first and last three sessions for both groups. Interestingly, in the first three sessions, the MBC was greater for the VI group, i.e., children used fewer blocks ( $p < 0.0001$ ). In contrast, when analysing the last three sessions, the MBC did not differ between either group.

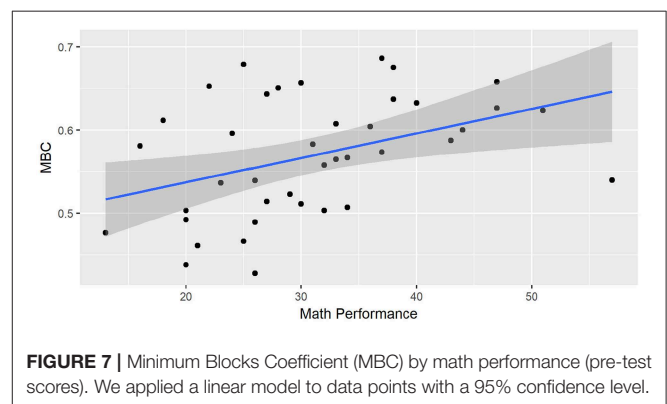
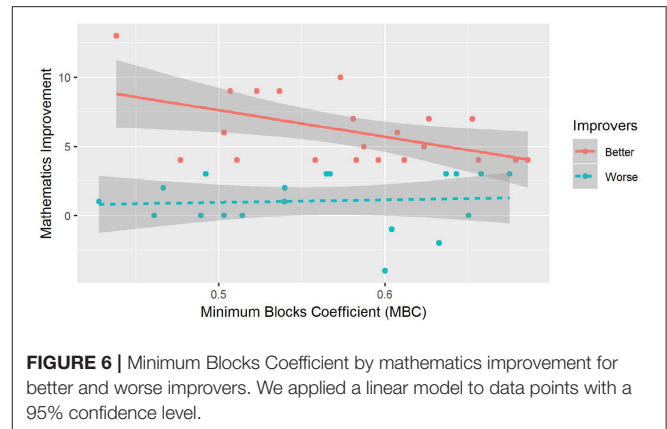
### 3.2.3. Minimum Blocks Coefficient and Mathematics Improvement

We explored the relationship between the number of blocks employed during the intervention (measured by MBC) and the amount of mathematical improvement (dScores: Post-Test scores – Pre-Test Scores) and found no correlation ( $p > 0.05$ ). Neither TI nor VI groups showed a significant correlation between MBC and dScore when analyzed separately ( $p > 0.05$ ).

Further, we decided to analyze the differences in the number of blocks employed comparing the performance of the Better and Worse Improvers. Thus, we divided all participants by the median of the dScore comprising two groups. The Better Improvers were the children with a dScore above the median, while the Worse Improvers were the ones whose dScore was below the median (see **Figure 6**). We found a significant negative correlation between MBC and dScores for the Better Improvers ( $cor = -0.50$ ,  $p = 0.021$ ), but not for the Worse Improvers. In conclusion, the children that had a greater improvement were the children using more blocks than the minimum blocks necessary to build the numbers required by the game. In contrast, we did not observe any change in the number of blocks used by the children who did not improve in mathematics.

### 3.2.4. Minimum Blocks Coefficient and Mathematics Performance

We were also interested in the relationship between the Minimum Blocks Coefficient (MBC) and mathematical



performance (Pre-Test scores). Analysis indicated that Pre-Test scores were positively correlated with the MBC ( $cor = 0.41$ ,  $p = 0.009$ ; see **Figure 7**). Children who had greater Pre-Test scores at the beginning of this study had the tendency to use less number of blocks during the game.

## 4. DISCUSSION

### 4.1. Impact of Manipulatives on Mathematical Learning

Our results indicate that the tangible manipulative group showed an advantage in mathematics scores after training compared to the control group. Our findings highlight the possibility of improving mathematical ability by practicing implicit number composition tasks assisted by tangible manipulatives.

We did not find significant differences either between the two types of manipulatives (virtual and tangible), or between virtual manipulatives and the control group when considering mathematical improvement tested by TEMA-3. It may be the case that virtual tangibles also have an impact in Post-Test scores, which was not observed due to the lack of statistical power of the present study.



## 4.2. Virtual and Tangible Manipulatives Led to Different Strategies in Number Composition

We analyzed children's behavior during our intervention to look for possible differential profiles in their evolution during training. Our tablet-based intervention allowed us to record the children's responses every time they submitted a block to compose a number. Our results enabled us to reflect on the role of specific actions performed by children affecting the learning process, and how learning could be influenced by the interactive properties of the blocks rendered as a representational assistance (Manches and O'Malley, 2016).

It was observed that the TI and VI groups significantly differed in the numbers of blocks used to compose a number. VI employed significantly fewer blocks compared with TI, showing that the different type of manipulatives could have led to different problem solving strategies. TI children opted to compose numbers using more varied combination of blocks, i.e., they used more number composition strategies. This suggests that the affordances of physical objects do trigger more diverse solutions (Manches and O'Malley, 2016), which have been advocated to prompt better learning experiences in numerosity knowledge (Alibali and Goldin-meadow, 1993; Chi et al., 1994; Siegler and Shipley, 1995) and specifically foster mastery of basic number combinations (Baroody and Tiilikainen, 2003; Sarana and Clements, 2009).

Our results are in accordance with Manches et al. (2010) results that found that children employed a significantly greater number of solutions when they used plastic blocks as manipulatives, comparing with a condition in which children were aided with a visual representation drawn on paper. For instance, it is easier to detect the "reversion" strategy (5-2, 2-5) when you can hold and displace objects representing these quantities (2 and 5). This finding supports the view that objects affordances implicitly carry information that could be relevant to reflect on abstract concepts, through conceptual metaphors. In our study, we compared tangible blocks (TI group) against virtual blocks (VI group). The use of virtual blocks allowed the children to drag, transform, and move blocks which allows a richer interaction compared to blocks drawn on paper. However, when compared to virtual blocks, tangible blocks enabled a more diverse combination of blocks to compose numbers as also observed elsewhere (Manches et al., 2010).

### 4.2.1. Strategies Evolution in Number Composition

When we analyzed strategies during training sessions we found that at the beginning of the training both groups employed more blocks to compose numbers with a tendency to diminish in the last sessions. This tendency to diminish may represent an approach to optimal performance (when the number is composed by the minimal quantity of possible blocks), probably reflecting learning toward increasing efficient and fastest strategies in number composition (Baroody and Dowker, 2003).

This is in line with the fact that composing and decomposing strategies becomes semiautomatic or automatic with effective

and faster answers to basic number combinations. Children may automatize some combinations of a number through practice, resulting in an association with their counting knowledge. This association encourages efficiency, preventing children from repeatedly practicing all the possible combinations (Baroody, 2006). In our study, children at the beginning started by practicing various combinations of numbers. For instance, in the first sessions to form the number 5 children might use several combinations as  $1+1+1+1+1$ ,  $2+2+1$ ,  $2+1+1+1$ , reflected by low MBC scores. Nevertheless, at the end of the training sessions children were able to answer more effectively, reflected by high MBC scores. For instance, to form the number 5 they answered with the block 5 or by adding just two blocks as  $2+3$  or  $4+1$ , which is quicker and more direct.

Analyses showed that the mean of blocks used in the first three sessions was significantly smaller for the VI group, whereas both groups employed the same number of blocks in the last three sessions. This suggests that besides the tendency of both groups to optimize responses, they presented a different profile in their evolution during training. Children who used tangible manipulatives had the tendency to use more blocks and showed a more pronounced decrease in the number of blocks used during the intervention compared to children who used virtual manipulatives. This finding may be connected to the observed improvement in maths scores (measured by TEMA-3) for the TI group. The number of combinations used in the TI may have contributed to achieving mastery in mathematical knowledge, since mastery in basic number composition is enriched by experiencing more varied possibilities (Markman, 1978; Bowerman, 1982; Karmiloff-Smith, 1992). In this study, physical object affordances offered the user a richer set of action possibilities, and most probably also a more comprehensive understanding of the phenomenon explored.

### 4.2.2. Strategies in Additive Composition Task and Mathematical Improvement

We did not find a correlation between the number of blocks employed by children and mathematical improvement in general (all children analyzed together). Nevertheless, when children were divided according to their improvement in mathematics (Post-Test – Pre-Test) after the intervention, it was observed that the greater improvement group showed a positive correlation between number of blocks employed and gain in mathematical knowledge, which was not found for the Worse Improvers.

Therefore, children who showed a greater improvement tended to use more blocks. This outcome may suggest that an optimal performance in number composition (understood as fewer pieces used to form a number equals better performance) would not necessarily lead to a better learning experience. Another hypothesis would be that children who do not already have this mastery in number combinations, i.e., efficient, fast and accurate responses, would benefit more from employing manipulatives to solve additive composition and this might be the case for the "Better Improvers." Children who improved at maths during training were the ones using more varied block combinations. This is connected to the fact that the use of a

greater variety of strategies can result in a better learning outcome (Markman, 1978; Bowerman, 1982; Karmiloff-Smith, 1992).

#### 4.2.3. Strategies in Additive Composition Task and Mathematics

Interestingly, a negative correlation was found between mathematical scores at the Pre-Test (how good the children were at the beginning of the study) and the number of blocks employed. That is, being better at mathematics at Pre-Test implied the use of fewer manipulative blocks, probably due to a better knowledge of retrieval strategies while composing numbers (Rathmell, 1978; Steinberg, 1985; Kilpatrick et al., 2001). Children who were good at maths at the beginning of the training will not necessarily use more strategies because they already have a deeper knowledge in number concept and composition. That is to say, children who have already learned basic combinations of numbers have the ability to use such knowledge to answer quickly and efficiently in a familiar and unfamiliar learning context (Baroody, 2006).

It may seem contradictory that children who obtained the best scores at TEMA-3 (better at mathematics at baseline) used fewer blocks whereas the Better Improvers tended to employ more. However, according to Sarama and Clements (2009), despite seeming paradoxical, those who are better at solving problems with objects, fingers or counting are less likely to persist in these strategies in the future—as already reported by Siegler (1993)—but this is because they trust their answers and therefore move toward more precise strategies based on the retrieval of number combinations, leaving behind what once served as a scaffolding.

These results also suggest that children who will benefit more from the use of manipulative blocks are the children who do not have already mastery in number combinations. The use of enhanced manipulatives may be more suitable for younger children who need to practice and automatize simple number combinations.

#### 4.3. Limitations

The present study has several limitations that should be considered when interpreting the results. It may lack statistical power since the number of participants in each group is small and for such reason, a larger confirmatory study is needed to strengthen the conclusions of the present study. The quasi-experimental design of the current study has more ecological validity (children were kept in their school groups), but it is susceptible to threats on internal validity compared to controlled experimental designs and for that reason we consider our results as exploratory and conclusions are drawn carefully.

#### 4.4. Conclusions

Current findings indicate that the use of tangible manipulatives had a positive impact on mathematical learning. We were able to observe interesting relationships between the level of mathematics and the kind of manipulative strategies chosen by the children when solving number composition tasks. Our results suggest that tangible manipulatives increase action possibilities and may also contribute to a deeper understanding of core mathematical concepts. Playing the game BrUNO with tangible manipulatives promotes meaningful practice of

more varied number combinations by encouraging children to focus on patterns and relationships in basic number combinations. In addition, we were able to observe how their responses pattern changed throughout the training leading to the use of less but efficient strategies in the last sessions which may reflect that they achieved mastery in doing such combinations. Thus, training in this basic combinations led to an improvement in mathematics and hopefully may lead children to effectively apply this knowledge in new and unfamiliar number combinations.

From an interaction design perspective (for more details regarding this research and perspective, see Marichal et al., 2017a), the most relevant observation is how the objects' affordances (i.e., the possibility to grasp physical objects or drag virtual ones) somehow shape and constrain users' strategies. In our study, tangible blocks meant a richer interaction, providing the opportunity to explore more number composition possibilities. This possibly led to an improvement in mathematical performance. Thus, depending on the learning task objective (context), we might take advantage of this phenomena, by choosing either tangible, virtual or mixed learning environments. The current study invites researchers to delve deeper in the exploration of the potential for designing interactive activities aimed at fostering learning of specific target content.

### ETHICS STATEMENT

All children that participated in this research had the informed consent form signed by their parents or legal guardians. The intervention current protocol was approved by the Local Research Ethical Committee of the Faculty of Psychology, and is in accordance with the 2008 Declaration of Helsinki.

### AUTHOR CONTRIBUTIONS

AP: substantial contributions to the conception or design of the work, analysis and interpretation of data for the work, drafting and revising it critically for important intellectual content. FG: drafting the work or revising it critically for important intellectual content, interpretation of data for the work. EB: substantial contributions to the conception and design of the work and data acquisition. BF: drafting the work or revising it critically for important intellectual content, analysis, and interpretation of data for the work. GS: substantial contributions to the design of the work. SM: substantial contributions to the design of the work, drafting and revising it critically for important intellectual content.

### FUNDING

This work was supported by Agencia Nacional de Investigación e Innovación (ANII); Fundación Ceibal, Espacio Interdisciplinario, and by Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje (CICEA), Universidad de la República, Uruguay; and by Fundação para a Ciência e a Tecnologia (FCT), I.P., through funding of project

mIDR (AAC 02/SAICT/- 2017, project 30347, cofunded by COMPETE/FEDER/FNR), and of the LASIGE Research Unit (UID/CEC/00408/2013), Portugal.

## ACKNOWLEDGMENTS

We would like to show our gratitude to the children and educators that participated in this work from Escuela Panamá,

Montevideo, Uruguay. We also want to thank our colleagues Rita Soria, María Pascale, Mariana Rodriguez, Leonardo Secco, Matías Correa, Leandro Fernández, Dilva Devita, Mariana Borges, Gonzalo Tejera, Alvaro Cabana, Fulvio Capurso, and Rodrigo López who provided insight and expertise that greatly assisted the research and design of CETA; and Catarina Tome-Pires and reviewers for suggestions and comments that greatly improved the manuscript.

## REFERENCES

- Alibali, M. W., and Goldin-meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: what the hands reveal about a child's state of mind. *Cogn. Psychol.* 25, 468–523.
- Anderson, S. E. (2003). The school district role in educational change: a review of the literature. *Int. Centre Educ. Change*. Available online at: <http://fcis.oise.utoronto.ca/~icec/workpaper2.pdf>
- Anstrom, T. (2006). Supporting students in mathematics through the use of manipulatives. *Am. Inst. Res.* 1–15. Available online at: <https://studylib.net/doc/18105288/supporting-students-in-mathematics-through-the-use-of-man>
- Antell, S. E., and Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Dev.* 54, 695–701. doi: 10.2307/1130057
- Antle, A. N., and Wise, A. F. (2013). Getting down to details: using theories of cognition and learning to inform tangible user interface design. *Interact. Comput.* 25, 1–20. doi: 10.1093/iwc/iws007
- 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*, Studies in Mathematical Thinking and Learning, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 1–33.
- Baroody, A. J. (2006). Why children have difficulties mastering the basic number combinations and how to help them. *Teach. Child. Math.* 13, 22–31. Available online at: <https://www.semanticscholar.org/paper/Why-Children-Have-Difficulties-Mastering-the-Basic-Baroody/753415c8d3fe2d6d3089bb0da0d9637dff02af32>
- Baroody, A. J., and Dowker, A. (eds.). (2003). *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Baroody, A. J., and Tiilikainen, S. H. (2003). “Two perspectives on addition development,” in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*, Studies in Mathematical Thinking and Learning, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 75–125.
- Bennett, S., Maton, K., and Kervin, L. (2008). The “digital natives” debate: a critical review of the evidence. *Br. J. Educ. Technol.* 39, 775–786. doi: 10.1111/j.1467-8535.2007.00793.x
- Bliss, S. (2006). Test of early mathematics ability-third edition. *J. Psychoeduc. Assess.* 24, 85–98. doi: 10.1177/0734282905282839
- Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: implications of developmental errors with causative verbs. *Quaderni Semant.* 3, 5–66.
- Canohi, 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
- Carpenter, T. P., and Moser, J. M. (1982). “The development of addition and subtraction problem-solving skills,” in *Addition and Subtraction: A Cognitive Perspective*, eds J. M. Moser, T. A. Romberg, and T. P. Carpenter (Hillsdale, NJ: Lawrence Erlbaum Associates), 9–24.
- Carpenter, T. P., and Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *J. Res. Math. Educ.* 15, 179–202.
- Chao, S.-J., Stigler, J. W., and Woodward, J. A. (2000). The effects of physical materials on kindergarten's learning of number concepts. *Cogn. Instruct.* 18, 285–316. doi: 10.1207/S1532690XCI1803\_1
- Chi, M. T., De Leeuw, N., Chiu, M.-H., and LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cogn. Sci.* 18, 439–477. doi: 10.1207/s15516709cog1803\_3
- Clements, D. H. (1999). Teaching length measurement: research challenges. *School Sci. Math.* 99, 5–11. doi: 10.1111/j.1949-8594.1999.tb17440.x
- Cuisenaire, G. (1968). *Les Nombres en Couleurs*. Bruxelles: Calozet.
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition*. New York, NY: Oxford University Press.
- Deng, Z., Gopinathan, S., and Lee, C. (2013). *Globalization and the Singapore Curriculum. From policy*. New York, NY: Springer.
- Dienes, Z. P. (1961). *Building Up Mathematics*. London: Hutchinson Educational.
- Feigenson, L., and Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Dev. Sci.* 6, 568–584. doi: 10.1111/1467-7687.00313
- 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
- Fuson, K. C. (1992). “Research on whole number addition and subtraction,” in *Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics*, ed D. A. Grouws (New York, NY: Macmillan Publishing Co, Inc.), 243–275.
- Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-R
- 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
- Gelman, R., and Gallistel, C. R. (1978). *The Child's Understanding of Number*. Cambridge: Harvard University Press.
- Groen, G., and Resnick, L. B. (1977). Can preschool children invent addition algorithms? *J. Educ. Psychol.* 69:645. doi: 10.1037//0022-0663.69.6.645
- Horn, M. (2012). *Topcode: Tangible Object Placement Codes*. Available online at: <http://users.eecs.northwestern.edu/~mhorn/topcodes/>
- Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10382–10385. doi: 10.1073/pnas.0812142106
- Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. London: MIT Press; Taylor and Francis.
- KAUFMAN, E. L. and LORD, M. W. (1949). The discrimination of visual number. *Am. J. Psychol.* 62, 498–525. doi: 10.2307/1418556
- Kilpatrick, J., Swafford, J., and Findell, B. (eds.). (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: National Academies Press.
- Klahr, D., Triona, L., Strand-Cary, M., and Siler, S. (2008). “Virtual vs. physical materials in early science instruction: transitioning to an autonomous tutor for experimental design,” in *Beyond Knowledge: The Legacy of Competence*, eds J. Zumbach, N. Schwartz, T. Seufert, and Kester L. (Dordrecht: Springer), 163–172.
- Levine, S. C., Jordan, N. C., and Huttenlocher, J. (1992). Development of calculation abilities in young children. *J. Exp. Child Psychol.* 53, 72–103. doi: 10.1016/S0022-0965(05)80005-0
- Manches, A. (2011). Digital manipulatives: tools to transform early learning experiences. *Int. J. Technol. Enhanc. Learn.* 3, 608–626. doi: 10.1504/IJTEL.2011.045451
- Manches, A., and O'Malley, C. (2016). The effects of physical manipulatives on children's numerical strategies. *Cogn. Instruct.* 34, 27–50. doi: 10.1080/07370008.2015.1124882



- Manches, A., O'Malley, C., and Benford, S. (2010). The role of physical representations in solving number problems: a comparison of young children's use of physical and virtual materials. *Comput. Educ.* 54, 622–640. doi: 10.1016/j.compedu.2009.09.023
- Marichal, S., Rosales, A., Sansone, G., Pires, A. C., Bakala, E., Sansone, G., et al. (2017a). "Ceta: designing mixed-reality tangible interaction to enhance mathematical learning," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '17* (New York, NY: ACM), 29:1–29:13.
- Marichal, S., Rosales, A., Sansone, G., Pires, A. C., Bakala, E., Perilli, F. G., et al. (2017b). "Ceta: open, affordable and portable mixed-reality environment for low-cost tablets," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '17* (New York, NY: ACM), 47:1–47:7.
- Markman, E. M. (1978). Empirical versus logical solutions to part-whole comparison problems concerning classes and collections. *Child Dev.* 49, 168–177. doi: 10.2307/1128605
- Martin, T., and Schwartz, D. L. (2005). Physically distributed learning: adapting and reinterpreting physical environments in the development of fraction concepts. *Cogn. Sci.* 29, 587–625. doi: 10.1207/s15516709cog0000\_15
- McGuire, P., Kinzie, M. B., and Berch, D. B. (2012). Developing number sense in pre-k with five-frames. *Early Childh. Educ. J.* 40, 213–222. doi: 10.1007/s10643-011-0479-4
- Miura, I. T., and Okamoto, Y. (2003). "Language supports for mathematics understanding and performance," in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*, Studies in Mathematical Thinking and Learning, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 229–242.
- Montessori, M. (1917). *The Advanced Montessori Method: Scientific Pedagogy as Applied to the Education of Children From Seven to Eleven Years*. London: W. Heinemann.
- Morin, J. E., and Franks, D. J. (2009). Why do some children have difficulty learning mathematics? Looking at language for answers. *Prevent. School Fail.* 54, 111–118. doi: 10.1080/10459880903217861
- Moyer-Packenham, P. S., and Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *IJVPLE* 4, 35–50. doi: 10.4018/ijvple.2013070103
- Oakes, J. M., and Feldman, H. A. (2001). Statistical power for nonequivalent pretest-posttest designs: the impact of change-score versus ancova models. *Eval. Rev.* 25, 3–28. doi: 10.1177/0193841X0102500101
- 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
- Rathmell, E. C. (1978). "Using thinking strategies to teach the basic facts," in *Developing Computational Skills*, ed M. N. Suydam (Reston: NCTM-Yearbook), 13–38.
- Sarama, J., and Clements, D. H. (2009). *Early Childhood Mathematics Education Research: Learning Trajectories for Young Children*. New York, NY: Routledge.
- Sarama, J., and Clements, D. H. (2016). "Physical and virtual manipulatives: what is 'concrete'?" in *International Perspectives on Teaching and Learning Mathematics With Virtual Manipulatives*, ed P. S. Moyer-Packenham (Cham: Springer International), 71–93.
- Schneider, B. A., Avivi-Reich, M., and Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Front. Psychol.* 6:474. doi: 10.3389/fpsyg.2015.00474
- Schwartz, D. L., Martin, T., and Pfaffman, J. (2005). How mathematics propels the development of physical knowledge. *J. Cogn. Dev.* 6, 65–88. doi: 10.1207/s15327647jcd0601\_5
- Siegler, R., and Jenkins, E. A. (2014). *How Children Discover New Strategies*. New York, NY: Psychology Press.
- Siegler, R. S. (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.
- Siegler, R. S. (1993). "Adaptive and nonadaptive characteristics of low-income children's mathematical strategy use," in *The Challenge in Mathematics and Science Education: Psychology's Response*, eds L. A. Penner, G. M. Batsche, H. M. Knoff, and D. L. Nelson (Washington, DC: American Psychological Association), 341–366.
- 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 Shipley, C. (1995). "Variation, selection, and cognitive change," in *Developing Cognitive Competence: New Approaches to Process Modeling*, eds T. J. Simon and G. S. Halford (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 31–76.
- Sophian, C., and McCorgray, P. (1994). Part – whole knowledge and early arithmetic problem solving. *Cogn. Instruct.* 12, 3–33. doi: 10.1207/s1532690xcil1201\_1
- Steffe, L. P., and Cobb, P. (1988). *Construction of Arithmetical Meanings and Strategies*. New York, NY: Springer-Verlag.
- Steinberg, R. M. (1985). Instruction on derived facts strategies in addition and subtraction. *J. Res. Math. Educ.* 16, 337–355. doi: 10.2307/749356
- Verschaffel, L., Greer, B., and De Corte, E. (2007). "Whole number concepts and operations," in *Second Handbook of Research on Mathematics Teaching and Learning*, ed F. Lester (Charlotte, NC: Information Age Publishing Inc.), 557–628.
- Wang, J., and Goldschmidt, P. (2003). Importance of middle school mathematics on high school students' mathematics achievement. *J. Educ. Res.* 97, 3–17. doi: 10.1080/00220670309596624
- Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322
- Wong, K. Y. (2009). *Mathematics Education: The Singapore Journey, Vol. 2*. Singapore: World Scientific Publishing Co.
- Wong, K. Y., and Lee, N. (2009). "Singapore education and mathematics curriculum," in *Mathematics Education: The Singapore Journey*, eds K. Y. Wong, P. Y. Lee, B. Kaur, P. Y. Foong, and S. F. Ng (World Scientific), 13–47.
- Yerushalmy, M. (2005). Functions of interactive visual representations in interactive mathematical textbooks. *Int. J. Comput. Math. Learn.* 10, 217–249. doi: 10.1007/s10758-005-0538-2

**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 Pires, González Perilli, Bakala, Fleisher, Sansone and Marichal. 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](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto: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