PROCESSING SYMBOLIC NUMERICAL INFORMATION AND ITS IMPLICATIONS FOR MATHEMATICS LEARNING

EDITED BY: Ricardo Moura, Julia Bahnmueller, Vitor Geraldi Haase,
Júlia Beatriz Lopes-Silva and Korbinian Moeller
PUBLISHED IN: Frontiers in Psychology and Frontiers in Education

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-158-6
DOI 10.3389/978-2-88976-158-6

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

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

PROCESSING SYMBOLIC NUMERICAL INFORMATION AND ITS IMPLICATIONS FOR MATHEMATICS LEARNING

Topic Editors:
Ricardo Moura, University of Brasilia, Brazil
Julia Bahnmueller, Loughborough University, United Kingdom
Vitor Geraldi Haase, Federal University of Minas Gerais, Brazil
Júlia Beatriz Lopes-Silva, Federal University of Minas Gerais, Brazil
Korbinian Moeller, Loughborough University, United Kingdom

Citation: Moura, R., Bahnmueller, J., Haase, V. G., Lopes-Silva, J. B., Moeller, K., eds. (2022). Processing Symbolic Numerical Information and its Implications for Mathematics Learning. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-158-6

Table of Contents

frontiers | Frontiers in Education

EDITORIAL
published: 25 April 2022
doi: 10.3389/feduc.2022.890804

Editorial: Processing Symbolic Numerical Information and Its Implications for Mathematics Learning

Ricardo Moura[1], Julia Bahnmueller[2], Vitor Geraldi Haase[3], Júlia Beatriz Lopes-Silva[3] and Korbinian Moeller[2]*

[1] Institute of Psychology, University of Brasília, Brasília, Brazil, [2] Department of Mathematics Education, Loughborough University, Loughborough, United Kingdom, [3] Department of Psychology, Federal University of Minas Gerais, Belo Horizonte, Brazil

Keywords: numerical processing, mathematics, place value, mathematics difficulties, cognitive development

Editorial on the Research Topic

Processing Symbolic Numerical Information and Its Implications for Mathematics Learning

OPEN ACCESS

Edited and reviewed by:
Douglas F. Kauffman,
Medical University of the Americas – Nevis, United States

****Correspondence:***
Ricardo Moura
ricardomoura@unb.br

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

Received: *06 March 2022*
Accepted: *16 March 2022*
Published: *25 April 2022*

Citation:
Moura R, Bahnmueller J, Haase VG, Lopes-Silva JB and Moeller K (2022) Editorial: Processing Symbolic Numerical Information and Its Implications for Mathematics Learning. Front. Educ. 7:890804. doi: 10.3389/feduc.2022.890804

INTRODUCTION

Among the first demands of enculturation on children's emerging numerical abilities is learning to transcode their initial representations of quantity into symbolic notations. It usually takes years of (formal) education before children master symbolic numerical representations including number words and digital-Arabic numerals as well as their place-value structuring and apply this successfully to perform arithmetic.

Understanding how symbolic and non-symbolic numerical representations relate to each other as well as predict numerical development may help to elucidate how children acquire numerical skills including arithmetic. In particular, as there is evidence suggesting that mastery of symbolic numerical representations—more so than non-symbolic ones—are building blocks for later arithmetic / mathematics performance.

The present Research Topic aimed at providing latest research results on processing of symbolic numerical information but also the trajectories in which processing of both symbolic and non-symbolic numerical information impacts numerical development. We collated 14 empirical studies focusing on (i) the semantic processing of symbolic numbers, (ii) the processing of symbolic numbers as a predictor of arithmetic / mathematics performance, and (iii) the understanding of the place-value structuring of symbolic numbers. These will be discussed in turn in the following.

SEMANTIC PROCESSING OF SYMBOLIC NUMBERS

One major question when it comes to the processing of number symbols is how they become associated with and under what circumstances they activate semantic magnitude information. The latter was investigated by Malykh et al. who observed that accuracy and speed of processing non-symbolic magnitudes developed differently from 1st to 9th grade, which might suggest different dependent variables reflect different underlying processes.

Finke et al. investigated cross-format activation of Arabic digits and number words in two ERP experiments with adult participants. They observed that number pairs seemed to be processed in two stages. At an early stage, number pairs presented in the same notation are integrated automatically without necessary access to semantic magnitude information. However, the latter is involved in a later second stage supporting cross-format integration of numbers.

In a further ERP study, van Hoogmoed et al. investigated the integration of numerical information presented in symbolic (Arabic digits) and non-symbolic (dot patterns) presentation formats. In line with the results of Finke et al.—but in contrast to ideas of an approximate number system underlying human numerical cognition—their results did not support the idea that children automatically activate semantic magnitude information when processing symbolic numbers.

When learning to understand symbolic numbers, a main challenge faced by children is to grasp the concept of the number zero and how to use this number in different numerical tasks and notations. In this context, Krajcsi et al. demonstrated that children understand verbal labels reflecting zero (e.g., nothing) and deal with empty sets even before they regard zero as a number.

Furthermore, Schmidt et al. investigated how early neuromotor experiences influence spatial associations for symbolic numbers. The authors considered two neuromuscular diseases characterized by progressive loss of motor abilities: spinal muscular atrophy (SMA, preventing any experiences of independent motoric exploration) and Duchenne muscular dystrophy (DMD, which compromises acquired experiences later in development). Results indicated that children with DMD exhibited typical spatial associations when processing symbolic numbers, while children with SMA exhibited no such or even reversed associations. These results corroborate the relevance of early sensorimotor experiences for children's numerical development.

When looking at numerical development in particular, it often is of specific interest if, and if so, which basic numerical skills—including the processing of symbolic numbers—predict later arithmetic / mathematics performance in what way. This question was addressed in another set of studies of our Research Topic.

PROCESSING SYMBOLIC NUMBERS AS A PREDICTOR OF ARITHMETIC / MATHEMATICS PERFORMANCE

It has been argued that mathematics achievement builds on more basic numerical skills with inconsistent findings regarding the relevance of processing non-symbolic vs. symbolic numerical information. In this context, Gloor et al. provided evidence that both Spontaneous Focusing on Numerosity as well as symbolic number skills longitudinally predicted mathematics achievement at the end of 1st grade. However, the contribution of symbolic numerical skills was observed to be more pronounced.

Furthermore, Chen et al. studied a sample of adolescents with congenital and acquired deafness to investigate associations between symbolic and non-symbolic magnitude processing and arithmetic performance. They observed significant associations between symbolic and non-symbolic processing and arithmetic performance even after controlling for demographic variables. Interestingly, however, number magnitude processing did not predict arithmetic performance in the group with congenital deafness indicating an influence of (hearing) language on numerical development.

Additionally, Räsänen et al. report on a cross-sectional large-scale study investigating the development of basic number processing skills (i.e., single-digit symbolic number comparison, digit-dot matching) and arithmetic fluency in children aged 9–15 years using data from the development of a screening tool for mathematics learning disabilities. They observed that girls performed better in tasks on basic number processing whereas boys performed better in tasks related to arithmetic fluency. This implies that gender should be considered when assessing mathematical learning disabilities.

Moreover, Abreu-Mendoza et al. evaluated the effectiveness of an intervention fostering non-symbolic proportional reasoning by means of Cuisenaire rods for improving symbolic fraction knowledge. Results indicated that the intervention significantly improved processing of non-symbolic continuous proportions, but did not significantly improve processing of discretized proportions such as symbolic fractions.

Furthermore, Vogel et al. investigated the association of mathematical abilities and the reversed numerical distance effect typically observed in order judgments of number sequences [i.e., faster RTs for sequences with small (2 3 4) compared to large distances (2 4 6)]. In an adult sample, they observed that not all individuals presented a significant reversed distance effect and that the association of order judgements with mathematical abilities was more pronounced for individuals who exhibited a significant reversed distance effect.

Finally, Loenneker et al. investigated the association of visuo-spatial and symbolic arithmetical skills in a sample of patients with Parkinson's Disease (PD) in different stages of the disease. Results indicated that the occurrence of arithmetic difficulties was predicted by attentional and visuo-spatial/constructional deficits even after controlling for clinical and sociodemographic confounds. Interestingly, patients' difficulties were mostly related to place value processing in calculation tasks, which highlights the relevance of evaluating this basic numerical skill in neuropsychological patients.

Importantly, place-value processing—reflecting knowledge of the structuring principle of symbolic numbers—is critical to numerical development. This was addressed in a final set of studies included in the Research Topic.

PLACE-VALUE UNDERSTANDING

Current research indicates that understanding the place-value structure of the Arabic number system is a challenge for children during the 1st years of primary school (e.g., Moura et al., 2015). At the same time, it is crucial for later mathematics achievement (e.g., Moeller et al., 2011). However, research into the theoretical underpinnings and developmental trajectories of place-value understanding as well as effective interventional approaches is currently limited.

Addressing these gaps, Herzog and Fritz-Stratmann link their recently proposed hierarchical model of place-value learning (Herzog et al., 2019) to number transcoding (i.e., writing numbers to dictation or reading digital-Arabic numbers aloud)—a commonly employed task used to index place-value understanding. The authors found that transcoding may indeed be a valid index for place-value understanding because 2nd and 3rd graders demonstrating more advanced levels of place-value understanding also performed better in writing Arabic numbers to dictation, especially for syntactically more complex numbers (e.g., including zeros).

Friedmann et al. further contribute to our theoretical knowledge of place-value understanding by presenting a single case study on the performance of a deaf participant who also presented specific deficits in transcoding multi-digit numbers. While word reading seemed to be preserved, reading and comprehension of multi-digit Arabic numbers was impaired with error patterns reflecting characteristics of the sign language used by the participant. The authors interpreted this as evidence for dissociated mechanisms for processing visual properties and the decimal structure of Arabic numbers.

Providing further insights regarding interventional approaches to foster place-value understanding, Yuan et al. evaluated in three studies whether and if so how children aged 4–6 years that had not yet entered school benefited from different approaches specifically designed to foster conceptual understanding of place-value in general and transcoding (i.e., reading multi-digit Arabic numbers) in particular. Based on their findings the authors argue that rather than traditional mathematical manipulatives (e.g., base-10-blocks, Abacus), shapes with a simple but easily graspable structure, or even exposure to pairs of multidigit Arabic numbers and their names, may be more effective for the acquisition of initial place-value understanding.

CONCLUDING REMARKS

As indicated by the wide range of topics addressed in this Research Topic, the relevance of symbolic processing for numerical cognition and its development is undeniable. However, even though it seems a very basic numerical skill, it nevertheless poses considerable challenges on learners especially with respect to acquiring understanding of the organizing place-value structuring principle of symbolic Arabic numbers. Taken together, this Research Topic indicates that understanding symbolic numbers (including their place-value structuring) seems key to successful numerical development.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Herzog, M., Ehlert, A., and Fritz, A. (2019). “Development of a sustainable place value understanding,” in *International Handbook of Mathematical Learning Difficulties*, eds A. Fritz, V. G. Haase, and P. Räsänen (Cham: Springer International Publishing), 561–579. doi: 10.1007/978-3-319-97148-3_33

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance-a longitudinal study on numerical development. *Res. Dev. Disabil.* 32, 1837–1851. doi: 10.1016/j.ridd.2011.03.012

Moura, R., Lopes-Silva, J. B., Vieira, L. R., Paiva, G. M., Prado, A. C. D. A., Wood, G., et al. (2015). From “five” to 5 for 5 minutes: Arabic number transcoding as a short, specific, and sensitive screening tool for mathematics learning difficulties. *Archiv. Clin. Neuropsychol.* 30, 88–98. doi: 10.1093/arclin/acu071

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 March 2021
doi: 10.3389/fpsyg.2021.589305

Developmental Changes in ANS Precision Across Grades 1–9: Different Patterns of Accuracy and Reaction Time

Sergey Malykh[1,2], Yulia Kuzmina[2] and Tatiana Tikhomirova[1,2]*

[1] Department of Psychology, Lomonosov Moscow State University, Moscow, Russia, [2] Psychological Institute of Russian Academy of Education, Moscow, Russia

The main aim of this study was to analyze the patterns of changes in Approximate Number Sense (ANS) precision from grade 1 (mean age: 7.84 years) to grade 9 (mean age: 15.82 years) in a sample of Russian schoolchildren. To fulfill this aim, the data from a longitudinal study of two cohorts of children were used. The first cohort was assessed at grades 1–5 (elementary school education plus the first year of secondary education), and the second cohort was assessed at grades 5–9 (secondary school education). ANS precision was assessed by accuracy and reaction time (RT) in a non-symbolic comparison test ("blue-yellow dots" test). The patterns of change were estimated via mixed-effect growth models. The results revealed that in the first cohort, the average accuracy increased from grade 1 to grade 5 following a non-linear pattern and that the rate of growth slowed after grade 3 (7–9 years old). The non-linear pattern of changes in the second cohort indicated that accuracy started to increase from grade 7 to grade 9 (13–15 years old), while there were no changes from grade 5 to grade 7. However, the RT in the non-symbolic comparison test decreased evenly from grade 1 to grade 7 (7–13 years old), and the rate of processing non-symbolic information tended to stabilize from grade 7 to grade 9. Moreover, the changes in the rate of processing non-symbolic information were not explained by the changes in general processing speed. The results also demonstrated that accuracy and RT were positively correlated across all grades. These results indicate that accuracy and the rate of non-symbolic processing reflect two different processes, namely, the maturation and development of a non-symbolic representation system.

Keywords: approximate number sense, non-symbolic comparison, speed-accuracy trade-off, general processing speed, numerical ratio effect

OPEN ACCESS

Edited by:
Ricardo Moura,
University of Brasilia, Brazil

Reviewed by:
Ricardo Scott,
University of Alicante, Spain
Camilo Gouet,
Pontificia Universidad Católica de Chile, Chile

****Correspondence:***
Sergey Malykh
malykhsb@mail.ru

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

Received: *05 August 2020*
Accepted: *03 March 2021*
Published: *24 March 2021*

Citation:
Malykh S, Kuzmina Y and Tikhomirova T (2021) Developmental Changes in ANS Precision Across Grades 1–9: Different Patterns of Accuracy and Reaction Time. Front. Psychol. 12:589305. doi: 10.3389/fpsyg.2021.589305

INTRODUCTION

Humans and other species are equipped with the ability to perceive and process numerical information without counting and using symbols (e.g., Cantlon and Brannon, 2007; Agrillo et al., 2009; Nieder and Dehaene, 2009). Particularly, people can rapidly estimate and compare sets of objects based on their numerosities to determine the largest one or detect changes in numerosity.

This ability can be supported by several systems of non-symbolic numerosity representations depending on the number of objects that should be perceived and the objects' separation.

The first system is subitizing, which is the ability to precisely estimate numerosity in cases in which the number of objects is less than 4 (e.g., Revkin et al., 2008). Subitizing is based on an object tracking system and requires attentional and working memory resources (Olivers and Watson, 2008; Vetter et al., 2008; Burr et al., 2010). If the number of objects is larger than 3–4 and the boundaries of the objects are distinct, the Approximate Number System (ANS) is activated to estimate numerosity (Burr and Ross, 2008; Viswanathan and Nieder, 2013). Numerous studies have also demonstrated that when the number of objects increases and they have high density, objects are likely to be perceived as an inseparable texture, and the third system – texture-density discrimination – is activated (e.g., Anobile et al., 2016; Pomè et al., 2019).

Among the three systems of non-symbolic numerosity estimation, the ANS is more often discussed regarding its relations with symbolic math skills and developmental changes across the preschool and school years (e.g., Halberda et al., 2008, 2012; Halberda and Feigenson, 2008). Various studies have identified the following two main features of the ANS: its imprecision and its rapidity. ANS imprecision manifests as the proportion of errors (PE) and the existence of the Numerical Distance Effect (NDE) or Numerical Ratio Effect (NRE). The NRE or NDE indicate that the PE in a non-symbolic comparison test increases when the sets are closer to each other in numerosity (Sasanguie et al., 2011; Lyons et al., 2015). The size effect manifests as growing imprecision in non-symbolic comparison and estimation when the numerosity of sets of objects increases, while the ratio between the two sets remains the same (Dehaene, 2001).

ANS Accuracy

To measure ANS, various tests are used. Among the most popular types of tests, in the numerosity comparison test, individuals compare two sets of objects (e.g., dots) and select the set that contains more objects (e.g., Gebuis and Reynvoet, 2012; Halberda et al., 2012; Norris and Castronovo, 2016). Different measures are used as indicators of ANS precision in non-symbolic comparison tests. In particular, accuracy (the proportion of correct answers) and the Weber fraction are the measures used in most studies (e.g., Halberda et al., 2012; Tosto et al., 2017). The Weber fraction reflects the smallest ratio between two sets of objects that can be reliably identified (Dietrich et al., 2015, 2016). In some cases, the NDE and NRE for accuracy can be calculated and are used as measures of ANS precision (e.g., Soltész et al., 2010; Lonnemann et al., 2011). Evidence suggests that accuracy-based measures are reliable and highly correlated with each other; thus, these measures can be used interchangeably (Inglis and Gilmore, 2014; Dietrich et al., 2016; Tosto et al., 2017). However, it has been shown that accuracy (proportion of correct answers) had the highest test-retest reliability among four possible measures of ANS precision (Inglis and Gilmore, 2014).

Usually, it is necessary to compare arrays of objects in a very short period. However, in different studies, the duration of the demonstration of the sets that must be compared varies. In particular, in the study by Halberda et al. (2008), the duration was 200 ms, whereas in the study by Smets et al. (2016), the duration was 1500 ms. Dietrich et al. (2016) manipulated the duration of the stimulus presentations (from 50 to 2400 ms) and demonstrated that ANS accuracy varied depending on the duration. It has been shown that the variance explained by the ratio between the two sets was higher under long reaction time (RT) conditions. As the authors noted, these results indicate that accuracy is more informative of the underlying numerosity representation under conditions with long presentation times.

Speed of Non-symbolic Processing

To consider the speed of processing non-symbolic information, the mean (or median) RT (in all tasks or correct answers) is used. Particularly, it has been postulated that individuals who are able to estimate numerosity faster have a more precise ANS (e.g., Mussolin et al., 2010; Halberda et al., 2012; De Smedt et al., 2013). Several studies used other measures based on the RT. Particularly, in the study by Vanbinst et al. (2012), the NDE was calculated based on the RT. It was assumed that the RT-based NDE indicated the effect of distance on the children's RT and that this effect was negative; hence, individuals who have higher ANS precision should demonstrate a lower NDE regarding RT.

However, RT-based measures are used less often than accuracy-based measures (Dietrich et al., 2016). Evidence suggests that RT-based measures (particularly the mean RT, NDE and NRE of RT) are not all correlated. In addition, accuracy-based measures are more informative regarding the underlying ANS acuity than RT-based measures (Dietrich et al., 2016). Particularly, it has been shown that there were no significant differences in RT between children with dyscalculia and children without such problems, whereas the differences in accuracy were significant (Piazza et al., 2010).

In addition to the low reliability of the measures based on RT, other methodological issues hinder the use of RT in ANS analyses. In particular, RT data usually violate the normal distribution assumption and demonstrate positive skewness. In addition, in some cases, in empirical data, influential values may distort the model fit (e.g., Baayen and Milin, 2010). As a normal distribution is an assumption of general linear models, some authors recommend applying different transformations to RT data to normalize the distribution (Whelan, 2008). However, other researchers do not recommend transforming RT data and demonstrate that transformation may not be beneficial or may distort the interpretation of the results (e.g., Ratcliff, 1993; Schramm and Rouder, 2019).

There are two different types of relationships between accuracy and RT and two different approaches to the interpretation of individual differences in RT (e.g., Dodonova and Dodonov, 2013). In the information-processing approach (Jensen, 2006), it is assumed that tasks are very simple and that errors are random. Hence, accuracy scores or PE do not significantly vary among individuals and cannot reflect individual differences in the ability to process non-symbolic information. In these cases, the RT is used to assess individuals' ability. It has been postulated that the RT is negatively correlated with ability and that individuals with higher ability (e.g., a more precise ANS) can perform tasks faster. Considering this assumption, it is expected that accuracy and RT should be negatively correlated in a non-symbolic comparison test.

In the education testing approach, the tasks might vary in their difficulties; thus, accuracy or PE can reflect individuals' ability. In such tests, the RT can also be used to measure individuals' ability, but the relationship between accuracy and RT might be more complex than that in the information-processing approach. When the RT and accuracy reflect the same construct, it is expected that the RT and accuracy could be negatively correlated. However, in some cases, individuals may prefer accuracy over speed and demonstrate a speed-accuracy trade-off (Ratcliff et al., 2015). In this situation, the RT and accuracy are positively correlated, complicating the interpretation of the results of tests based on RT measures only.

The association between RT and accuracy in complex tests might vary depending on the task difficulty and individuals' ability. Evidence suggests that in easy tasks, the RT and accuracy are negatively correlated, whereas in more difficult tasks, the RT and accuracy are positively correlated (e.g., Neubauer, 1990; Dodonova and Dodonov, 2013). Particularly, it was demonstrated that in the Raven test, there was a difference in RT, but not in accuracy, in response to easy items between high-ability and low-ability individuals. Concurrently, high-ability individuals differed from low-ability individuals in terms of the rate of change in accuracy in response to more difficult items, but no differences in RT changes were observed (Dodonova and Dodonov, 2013).

Regarding the ANS, evidence suggests that the RT and accuracy are positively correlated; accordingly, a speed-accuracy trade-off has been found (e.g., Dietrich et al., 2016). Dietrich et al. (2016) noted that if participants showed a speed-accuracy trade-off, the accuracy and RT provided controversial information regarding the ability to process numerosity information in a non-symbolic format. However, in some studies, a negative correlation was found between ANS accuracy and the RT (e.g., Soltész et al., 2010; Libertus et al., 2013). Hence, the relationships between accuracy and RT in a non-symbolic comparison test might change depending on the sample or test difficulty.

Although it has been shown that accuracy is more informative regarding ANS precision than RT, the RT can reflect an important aspect of non-symbolic representation. Particularly, the RT was found to explain 5–8% of the variance in math performance in addition to the variance explained by ANS accuracy (Libertus et al., 2011). Moreover, it has been shown that the speed of different tests of non-symbolic comparison formed a separate latent factor distinct from accuracy (Soltész et al., 2010). Some authors claimed that it is necessary to consider both accuracy and the RT in assessing the characteristics of cognitive processes (e.g., Ratcliff et al., 2015). In summary, previous findings revealed that accuracy and RT might reflect different processes and cannot be used interchangeably as measures of ANS precision (Dietrich et al., 2016). Hence, investigations of developmental changes in ANS precision require an estimation of changes in both accuracy and RT.

Developmental Changes in ANS Accuracy and RT

Some evidence suggests that ANS precision increases throughout development. Most studies investigating developmental changes in the ANS have been performed based on changes in accuracy (e.g., Odic, 2018; Tikhomirova et al., 2019; Kuzmina et al., 2020) or Weber fraction (Halberda and Feigenson, 2008; Halberda et al., 2012). Weber fraction was found to decrease (e.g., Odic et al., 2013), whereas accuracy was found to increase across ages (e.g., Tikhomirova et al., 2019).

Although the hypothesis that ANS precision in adults is higher than that in children has been confirmed in various cross-sectional studies, longitudinal studies have cast doubt regarding the growth in ANS precision as a general phenomenon. Particularly, it has been demonstrated that growth in ANS precision slows by the end of elementary school (Tikhomirova et al., 2019). Latent growth models revealed that a significant proportion of pupils did not demonstrate growth in ANS accuracy (Tikhomirova et al., 2019). In addition, the increases in accuracy in non-symbolic comparison were found to be significant only among pupils with a high level of fluid intelligence or processing speed (PS) (Kuzmina et al., 2020).

Evidence suggests that the RT in non-symbolic comparison tests also changes across development. It has been demonstrated that adults have lower RTs in non-symbolic comparison tests than children (Halberda et al., 2012). In particular, Halberda and colleagues revealed that the RT rapidly decreased from the ages of 11 to 16 years, and then, the rate of change slowed, while accuracy continued to improve from the ages of 16 to 30 years (Halberda et al., 2012).

It has also been postulated that the development of non-symbolic representation precision is related to decreasing NDE or NRE (for a discussion, see Lyons et al., 2015). Particularly, adults demonstrated a lower distance effect than children (Halberda and Feigenson, 2008; Holloway and Ansari, 2008). Neurophysiological evidence further suggests that differences exist in the distance effect between adults and children. The amount of activation in the intraparietal sulcus (IPS) has been found to decrease as the numerical distance increases (Pinel et al., 2001). Ansari and Dhital (2006) demonstrated that adult participants exhibited greater effects of numerical distance in the left IPS than children. The authors suggested that these differences were related to developmental shifts from more controlled to more automatic processing of the numerical magnitude (Ansari and Dhital, 2006). It is possible that the development of ANS precision might involve changes in both accuracy and RT, reflecting improvement in general PS.

Development of General PS

A large body of evidence suggests that general PS increases across development (e.g., Kail, 2000; Kail and Ferrer, 2007; Nettelbeck and Burns, 2010; Coyle et al., 2011). It has been demonstrated that exponential and quadratic models of changes in general PS fit the data better than other models (e.g., linear, hyperbolic, and inverse regression models). It has been hypothesized that the patterns of changes in PS (linear increase with non-linear decrease) are consistent with the patterns of quadratic changes in physical growth in childhood and adolescence (Kail and Ferrer, 2007). Improvement in general PS is associated with the process of myelination and white matter integrity across childhood (Mabbott et al., 2006; Scantlebury et al., 2014; Chevalier et al., 2015; Chopra et al., 2018).

Alternative theories regarding the developmental trends in PS and its relationships with the development of other cognitive functions have been developed. The global trend hypothesis posits that all cognitive, motor and perceptual processes develop at the same rate (e.g., Hale, 1990; Kail, 1991). Kail (2000) suggested that general mechanisms limit the speed of processing information regardless of the task specificity. Particularly, it has been shown that PS in tasks, such as mental addition, mental rotation and simple motor skills, improved across development at a common rate according to an exponential function (Kail, 1991).

The alternative local trend hypothesis posits that all components of information processing develop at different rates (Bisanz et al., 1979). It has also been hypothesized that the rate of change in the speed of cognitive processes might be domain-specific, whereas within one domain, all components develop at a common rate (Kail and Miller, 2006). For example, it has been shown that children aged between 9 and 14 years have a faster PS in language tasks than non-language tasks. However, the rate of change in the PS of non-language tasks was faster than that in language tasks (Kail and Miller, 2006).

Improvement in general PS affects further improvement in other cognitive functions, such as working memory, intelligence, inhibition, math skills and reasoning ability (e.g., Fry and Hale, 1996; Kail et al., 2016). In particular, the following development cascade has been demonstrated: the general PS affects further improvement in working memory and intelligence, which, in turn, might affect improvement in general PS (Fry and Hale, 1996; Nettelbeck and Burns, 2010). It has also been shown that improvement in general PS partially explains the changes in general intelligence and accuracy of non-symbolic representation (Pezzuti et al., 2019; Kuzmina et al., 2020). However, the extent to which the changes in non-symbolic PS are explained by the development of general PS is unknown. From the perspective of the global trend hypothesis, age-related changes in an individual's speed in a non-symbolic comparison test should be explained by age-related changes in general PS. The local trend hypothesis implies that the patterns and rates of change in general and non-symbolic PS might differ.

Current Study

Considering the complex relationships among accuracy, RT and ability level, we hypothesized that developmental changes in the ANS should be analyzed while considering developmental changes in both accuracy and RT. Moreover, in previous studies, a speed-accuracy trade-off was found in non-symbolic comparison tests, but the developmental changes in the relationship between accuracy and RT in non-symbolic comparison tests are unknown. As the NRE is a core feature of non-symbolic representation, it is crucial to estimate developmental changes by considering the NRE.

In summary, our research aims are as follows:

(1) To assess developmental changes in accuracy and RT in non-symbolic representation across the school years,
(2) To assess the extent to which developmental changes in accuracy and RT vary depending on the ratio between compared arrays,
(3) To estimate the developmental relationships between accuracy and RT in a non-symbolic comparison test, and
(4) To estimate the extent to which the changes in general PS may explain the changes in accuracy and RT in a non-symbolic comparison test.

MATERIALS AND METHODS

Sample

This study was conducted using data collected for an ongoing longitudinal project named the "Cross-cultural Longitudinal Analysis of Student Success" (CLASS) project. For the aim of this study, two cohorts of schoolchildren studying in one school in the Moscow region were tested. This school was a state school with no selection of pupils.

The first cohort was tested from grade 1 to grade 5. In total, 313 pupils were tested, but some pupils participated less than three times due to illness and absence from school on the date of testing. As at least three time points are necessary to carefully estimate developmental trajectories and development relationships (e.g., Duncan and Duncan, 2009; Curran et al., 2010), the data of the schoolchildren who participated once or twice were removed from the analysis. The patterns of missing data in the sample were tested, and the missing completely at random (MCAR) assumption was confirmed by Little's MCAR test (1988) (Little, 1988). This test was non-significant (chi-square distance = 69.49, df = 64, $p = 0.30$), indicating that the MCAR assumption held. Since the MCAR assumption held and the sample size was sufficient, list-wise deletion can be applied to obtain adequate parameter estimates (Coertjens et al., 2017). The remaining sample consisted of 260 pupils (49% were girls, the mean age in grade 1 was 7.84, range 6.81–8.86), 17% of the pupils participated three times, 44% of the pupils participated four times, and 39% of the pupils participated five times.

The second cohort was tested from grade 5 to grade 9. The initial sample consisted of 246 pupils. Meanwhile, some pupils participated in the survey less than three times. To assess the growth trajectories more precisely, we analyzed the data of the pupils who participated at least three times. The patterns of missing data in the sample were tested, and the MCAR assumption was confirmed by Little's MCAR test (1988) (Little, 1988). This test was non-significant (chi-square distance = 57.77, df = 59, $p = 0.52$), indicating that the MCAR assumption held. The final sample consisted of 210 pupils (52% were girls, the mean age in grade 5 was 11.82 years, range 10.54–12.57), 11% of the pupils participated three times, 38% of the pupils participated four times, and 51% of the pupils participated five times.

This study received approval from the Ethics Committee of the Psychological Institute of the Russian Academy of Education.

Parental informed and written consent was obtained prior to the data collection. Consent was obtained from the children orally.

Procedures and Instruments

The pupils were assessed at the end of the academic year (April–May). All participants were tested in quiet settings within their school facilities by trained experimenters. All experimenters strictly used the same protocol and instructions for the test administration across all measurements. The pupils completed non-symbolic comparison and general PS tests in the computer form. The experiment was performed in a computer classroom in groups of 14–15 pupils. Each pupil sat in front of an individual monitor screen approximately 60 cm from the screen and performed the experiment independently. Each computer had a 17" LCD display with a resolution of 1440–900 pixels and a refresh rate of 60 Hz.

ANS

A non-symbolic comparison test was used to estimate ANS precision at each time point. The participants were presented arrays of yellow and blue dots in an intermixed format and varying in size and number. The task required the participants to judge whether the array contained more yellow or blue dots by pressing the corresponding keys on the keyboard (for the yellow dots, the participants pressed the "ж" key, corresponding to the ":" key on a QWERTY keyboard; for the blue dots, the participants pressed the "c" key, corresponding to the "c" key on a QWERTY keyboard). The following instructions were provided: "In this test, a number of circles will flash on the screen for less than half a second. The circles differ in size, and each circle is either yellow or blue. Your job is to judge whether you see more yellow or more blue circles flashing on the screen. If you think that there are more YELLOW circles, press "Y" on your keyboard. If you think that there are more BLUE circles, press "B" on your keyboard. Your decision must be based on the number of circles and not the sizes of the circles. In some trials, it may be difficult to judge. Don't worry! Let your "number sense" guide you and go with your instinct. This test should take less than 10 min. You should try to complete the test in one session. However, if you prefer, you will be able to take a break at certain places in the test where you will see a "come back later" button. Remember, we are measuring speed and accuracy, so please respond as quickly as you can. Press the SPACE BAR to start."

The stimuli included 150 static pictures, with the arrays of yellow and blue dots varying between 5 and 21 dots of each color and the ratios of the arrays of the two colors falling between 0.30 and 0.87. All trials can be divided into the following five ratio bins: 0.30–0.60 (23 trials), 0.61–0.75 (33 trials), 0.76–0.80 (29 trials), 0.81–0.84 (35 trials), and 0.85–0.87 (30 trials).

The presentation order was the same for all participants. Each stimulus flashed on the screen for 400 ms, and the maximum response time was 8 s. If no answer was given during this time, the answer was recorded as incorrect, and a message appeared on the screen to encourage the participant to press the space bar to continue to the next trial. The message disappeared after 20 s, and the next trial was displayed only after pressing the space bar. The task included a practice trial with two items and an option to repeat the practice trial.

In each trial, the cumulative area of the set containing more dots was larger than the cumulative area of the other set. The ratio of the cumulative areas of the two sets (the smallest area divided by the largest area) ranged between 0.30 and 0.87. In all trials, the average size of the yellow dots was equal to the average size of the blue dots.

To assess ANS precision, the following two measures were calculated: accuracy (proportion of correct answers) and RT (mean RT of the correct responses).

General PS

Processing speed was measured via modification of an RT test (Deary et al., 2001). In this version, the numbers 1, 2, 3, and 4 appeared 10 times each in a randomized order at random intervals between 1 and 3 s. The interval of 1 s was repeated 14 times, and intervals of 2 and 3 s between the presentations were repeated 13 times each. The task consisted of pressing the key corresponding to the number appearing on the screen as fast and accurately as possible. One series of numbers was used for all participants. The task started with instructions and a practice trial consisting of 6 items. The following instructions were provided: "This test should take only 2 or 3 min. You will need to complete the test in one go as there is no "come back later" option. We want to measure your speed, so please respond as quickly and as accurately as you can. You are going to see the numbers 1, 2, 3, and 4 flashing in the middle of the screen one at a time. Each time a number appears, press the matching key at the top of your keyboard as quickly as you can. To respond rapidly, you should position your left fingers on the keys "1" and "2" and your right fingers on the keys "3" and "4" as shown in the picture. Remember to only use the number keys at the top of the keyboard." The practice trial could have been repeated. The time out for responses was 8 s. If no response was given during this time, the next trial followed. The mean RT of the correct responses was calculated as an indicator of PS. Lower RTs corresponded to higher general PS.

Statistical Approach

First, we examined the accuracy and RT of the correct answers in each cohort and grade. To account for the non-symbolic comparison ratio dependence, we inspected the accuracy and RT in the following five ratio bins: 0.30–0.60 (23 trials), 0.61–0.75 (33 trials), 0.76–0.80 (29 trials), 0.81–0.84 (35 trials), and 0.85–0.87 (30 trials). The ratio was calculated as the smallest number divided by the largest number; thus, a larger ratio was associated with a decreasing distance between two numerosities that need to be compared. Next, the correlations between the accuracy and RT of the correct answers were estimated in each grade. To estimate the significance of the differences between the smallest and largest ratio bins in accuracy and RT, a paired-samples t-test was used.

To estimate the average and individual growth trajectories of non-symbolic representation, we used the mixed-effect growth approach (ME approach). The ME approach considers repeated measures that change over time "nested" in individuals. This approach allows researchers to estimate the average trajectory of the entire sample and individual-specific deviations from the average trajectory of each person. According to this framework, the intercept and the slope may vary across individuals, and this heterogeneity is described by the variance in the intercept and the slope.

We tested several models and used the likelihood ratio test (LR test) to choose the best-fitting model of the accuracy and mean RT of the correct answers as outcomes. For each cohort and outcome, several models were tested. The analysis started with testing the intercept-only model. This model estimates the intercept and between- and within-individual variance. The proportion of between-individual variance to the total variance (ICC) obtained from this model reflects the stability of outcomes across time. Higher ICC values correspond to greater between-individual variability and smaller within-individual variability (or greater time stability).

In several subsequent models, different patterns of changes were tested (linear changes and quadratic changes). We also tested random slope models and compared these models with a fixed slope model. A random slope model implies that the slope of the time variable varies across individuals. Hence, there were significant differences between individuals in the rate of change in ANS precision across the grades. In this model, the variance in the slope of the time variable and the covariance between the individual deviation of the slope and the intercept were estimated. To investigate the relationships between the changes in accuracy and RT, for each individual, the predicted growth in accuracy and RT were calculated, and the correlations between these measures were estimated.

Next, to estimate the extent to which the changes in general PS explain the changes in RT and accuracy in the non-symbolic comparison, the general PS was added to the model as a predictor of ANS, and RT and accuracy were the outcomes. If general PS explains the changes in ANS RT and accuracy, the coefficients of the "time" variable decrease or become insignificant. Finally, to compare the developmental patterns of general and non-symbolic PS, we estimated and compared the growth trajectories of general PS and non-symbolic PS.

TABLE 1 | Descriptive statistics for accuracy and RT for all trials and for five ratio bins.

Grade	Accuracy (proportion of correct answers)											
	All		Bin1 0.30–0.60		Bin2 0.61–0.75		Bin3 0.76–0.80		Bin4 0.81–0.84		Bin5 0.85–0.87	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Cohort 1												
1	0.63	0.09	0.75	0.17	0.66	0.12	0.60	0.11	0.58	0.11	0.58	0.10
2	0.65	0.08	0.79	0.15	0.68	0.12	0.63	0.12	0.60	0.10	0.61	0.10
3	0.68	0.08	0.82	0.12	0.71	0.11	0.65	0.11	0.62	0.10	0.62	0.10
4	0.69	0.08	0.85	0.13	0.73	0.12	0.67	0.11	0.63	0.10	0.62	0.11
5	0.69	0.08	0.85	0.12	0.72	0.11	0.67	0.10	0.63	0.09	0.63	0.11
Cohort 2												
5	0.67	0.09	0.83	0.14	0.70	0.12	0.65	0.12	0.62	0.11	0.61	0.11
6	0.69	0.09	0.84	0.15	0.72	0.13	0.66	0.12	0.63	0.09	0.64	0.10
7	0.69	0.09	0.84	0.14	0.73	0.13	0.66	0.11	0.65	0.10	0.63	0.10
8	0.73	0.08	0.89	0.12	0.76	0.12	0.70	0.10	0.66	0.10	0.67	0.10
9	0.75	0.07	0.92	0.09	0.80	0.11	0.72	0.09	0.68	0.09	0.67	0.10
Grade	**RT for correct answers (sec.)**											
	All		Bin1 0.30–0.60		Bin2 0.61–0.75		Bin3 0.76–0.80		Bin4 0.81–0.84		Bin5 0.85–0.87	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Cohort 1												
1	1.51	0.51	1.51	0.46	1.49	0.53	1.51	0.56	1.54	0.58	1.51	0.59
2	1.37	0.39	1.38	0.39	1.35	0.44	1.37	0.43	1.37	0.44	1.40	0.46
3	1.23	0.30	1.20	0.28	1.22	0.35	1.23	0.33	1.25	0.36	1.25	0.36
4	1.10	0.27	1.07	0.25	1.10	0.30	1.09	0.30	1.13	0.31	1.13	0.32
5	1.01	0.26	0.96	0.21	1.01	0.28	1.01	0.28	1.02	0.29	1.03	0.28
Cohort 2												
5	1.04	0.25	1.03	0.26	1.04	0.28	1.02	0.28	1.06	0.30	1.06	0.30
6	0.93	0.23	0.90	0.22	0.92	0.25	0.94	0.25	0.95	0.27	0.95	0.26
7	0.85	0.22	0.82	0.19	0.85	0.26	0.85	0.23	0.88	0.26	0.87	0.27
8	0.88	0.19	0.81	0.16	0.90	0.23	0.92	0.24	0.94	0.25	0.92	0.24
9	0.89	0.17	0.81	0.14	0.89	0.20	0.90	0.20	0.93	0.21	0.93	0.21

TABLE 2 | Correlations between accuracy and mean RT for correct answers.

Grade	Correlation between accuracy and RT for correct answers
Cohort 1	
Grade 1	0.53***
Grade 2	0.45***
Grade 3	0.32***
Grade 4	0.52***
Grade 5	0.42***
Cohort 2	
Grade 5	0.50***
Grade 6	0.47***
Grade 7	0.68***
Grade 8	0.55***
Grade 9	0.39***

****p < 0.001.*

RESULTS

Descriptive Statistics

The descriptive statistics of the accuracy and RT in the non-symbolic comparison test in the whole test and five ratio bins are presented in **Table 1**. The results revealed that across all grades, the highest accuracy was obtained with the smallest proportion (for ratio 0.30–0.60).

Significant positive correlations were observed between the RT of the correct answers and accuracy in grades 1–5 in Cohort 1 and grades 5–9 in Cohort 2 (**Table 2**). Hence, a speed-accuracy trade-off was found in each grade, although the values of the correlations varied across the grades.

The descriptive statistics of general PS are presented in **Supplementary Material Table 3**.

Estimation of the Ratio Dependence in the ANS Accuracy and RT

A paired-samples *t*-test was conducted to estimate the significance of the differences in accuracy between the smallest ratio bin (0.30–0.60) and the largest ratio bin (0.85–0.87). The analysis revealed that the difference in accuracy between the smallest and largest ratio bins was significant in both cohorts across all time points (**Table 3**). The analysis also demonstrated that in both cohorts, the effect size of the difference between the two ratio bins increased across time.

The analysis of the difference in the ANS RT between the ratio bins revealed that the difference was insignificant in grades 1–2 in Cohort 1 and grade 5 in Cohort 2 (**Table 3**). These results revealed that ANS precision varied depending on the ratio

TABLE 3 | Results of paired-sample *t*-test for differences in ANS accuracy and RT between the smallest and the largest ratio bins.

Grade	Bin1 (ratio 0.30–0.60)		Bin5 (ratio 0.85–0.87)		Mean difference (95% CI)	t	df	Effect size (Cohen's *d*)
	M	SD	M	SD				
Accuracy (Cohort 1)								
1	0.75	0.17	0.58	0.10	0.17 (0.15; 0.19)	15.61***	179	1.23
2	0.79	0.15	0.61	0.10	0.18 (0.16; 0.20)	18.24***	218	1.43
3	0.82	0.12	0.62	0.10	0.21 (0.19; 0.22)	25.99***	227	1.87
4	0.85	0.13	0.62	0.11	0.22 (0.21; 0.23)	26.61***	238	1.86
5	0.85	0.12	0.63	0.11	0.22 (0.21; 0.24)	29.72***	232	1.95
Accuracy (Cohort 2)								
5	0.83	0.14	0.61	0.11	0.21 (0.19; 0.23)	24.16***	186	1.68
6	0.84	0.15	0.64	0.10	0.20 (0.18; 0.20)	18.00***	166	1.60
7	0.84	0.14	0.63	0.10	0.20 (0.18; 0.22)	20.48***	182	1.63
8	0.89	0.12	0.67	0.10	0.22 (0.21; 0.24)	27.90***	198	2.05
9	0.92	0.09	0.67	0.10	0.25 (0.23; 0.26)	29.59***	189	2.52
RT (sec.) (Cohort 1)								
1	1.51	0.46	1.51	0.59	−0.00 (−0.05; 0.05)	−0.03	179	−0.002
2	1.38	0.39	1.40	0.46	−0.01 (−0.05; 0.03)	−0.47	218	−0.02
3	1.20	0.28	1.25	0.36	−0.05 (−0.08; −0.02)	−3.02**	227	−0.16
4	1.07	0.25	1.13	0.32	−0.06 (−0.09; −0.03)	−4.60***	238	−0.21
5	0.96	0.21	1.03	0.28	−0.07 (−0.09; −0.04)	−5.36***	232	−0.27
RT (sec.) (Cohort 2)								
5	1.03	0.26	1.06	0.30	−0.02 (−0.06; 0.01)	−1.27	186	−0.08
6	0.90	0.22	0.95	0.26	−0.05 (−0.08; −0.02)	−3.16**	166	−0.19
7	0.82	0.19	0.87	0.27	−0.04 (−0.07; −0.02)	−3.28**	182	−0.19
8	0.81	0.16	0.92	0.24	−0.08 (−0.10; −0.06)	−7.40***	198	−0.41
9	0.81	0.14	0.93	0.21	−0.12 (−0.14; −0.09)	−10.56***	189	−0.65

****p < 0.001, **p < 0.01.*

between the two compared arrays, although these differences mostly manifested in accuracy rather than RT.

Developmental Changes in ANS Accuracy

The growth trajectories of ANS accuracy measured by the non-symbolic comparison test were estimated in Cohort 1 (grades 1–5) and Cohort 2 (grades 5–9) separately. The results of Cohort 1 are presented in **Table 4**.

The results of the ME growth model of Cohort 1 revealed that the model with non-linear changes and a random slope fit the data better than the models with linear changes or a fixed slope. The values of the coefficients of the variables "time" and "time^2" demonstrated that ANS accuracy increased from grade 1 to grade 5, but growth slowed after grade 3. The results of post-estimation revealed that there was no difference in average predicted accuracy between grades 3 – 5 (**Supplementary Table 1**). The covariance between the intercept and slope at the individual level was significant and negative, indicating that the pupils who had a higher level of accuracy at grade 1 demonstrated less growth (**Supplementary Figure 1**).

The results of the changes in accuracy in Cohort 2 (grades 5–9) are presented in **Table 5**. The results of the pupils in grades 5–9 revealed that the model with non-linear changes and a random slope fit the data better than the model with linear changes and a fixed slope (**Supplementary Figure 2**). The results of post-estimation indicated that predicted average accuracy did not increase from grade 5 to grade 7 but increased later (grades 8–9) (**Supplementary Table 2**).

The comparison of the average accuracy in grade 5 in both cohorts revealed that there is no difference in accuracy in grade 5 between the two cohorts. The analysis of the average growth trajectories in grades 1–5 and 5–9 indicated that accuracy was relatively stable from grade 3 to grade 7 (**Figure 1**).

TABLE 4 | Cohort 1: results of ME growth models for changes in ANS accuracy from grade 1 to grade 5.

Variables	Baseline	Model 1	Model 2	Model 3
	Intercept-only	Linear growth	Non-linear growth	Model with random slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	0.67*** (0.004)	0.63*** (0.005)	0.62*** (0.006)	0.62*** (0.006)
Time		0.02*** (0.001)	0.036*** (0.005)	0.037*** (0.004)
Time^2			−0.005*** (0.001)	−0.005*** (0.001)
Random effect				
Intercept variance	0.003	0.003	0.003	0.004
Residuals	0.005	0.004	0.004	0.004
Slope variance (time)				0.0002
Covariance between intercept and slope				−0.0004
Log-likelihood	11191.9	1255.53	1262.74	1270.79
LR test (Δ df)		127.27*** (1)	14.42*** (1)	16.11*** (2)
ICC	0.36			

$^{***}p < 0.001$.

TABLE 5 | Cohort 2: results of ME growth model for changes in ANS accuracy from grade 5 to grade 9.

Variables	Baseline	Model 1	Model 2	Model 3
	Intercept-only	Linear growth	Non-linear growth	Model with random slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	0.71*** (0.004)	0.67*** (0.005)	0.68*** (0.006)	0.68*** (0.006)
Time		0.018*** (0.001)	0.005 (0.005)	0.005 (0.006)
Time^2			0.003** (0.001)	0.003** (0.001)
Random effect				
Intercept variance	0.005	0.004	0.004	0.005
Residuals	0.003	0.003	0.003	0.003
Slope variance (time)				0.0002
Covariance between intercept and slope				−0.0006
Log-likelihood	1005.68	1075.70	1079.42	1088.91
LR test (Δ df)		140.03*** (1)	7.44** (1)	20.31*** (2)
ICC	0.39			

$^{***}p < 0.001$, $^{**}p < 0.01$.

Changes in the Speed of Non-symbolic Processing

Furthermore, we estimated the patterns of the changes in the speed of non-symbolic processing as measured by the RT in the non-symbolic comparison task. The results of the ME growth model of the RT of the correct answers in Cohort 1 are demonstrated in **Table 6**.

The analysis revealed that the RT of the correct answers decreased from grade 1 to grade 5 according to a linear pattern as the model with non-linear changes did not fit the data better than the model with linear changes. The model with a random slope fit the data better than the model with a fixed slope. The covariance between the individual intercept and slope was negative, indicating that individuals with a larger RT in grade 1 had greater changes in RT (**Supplementary Figure 3**).

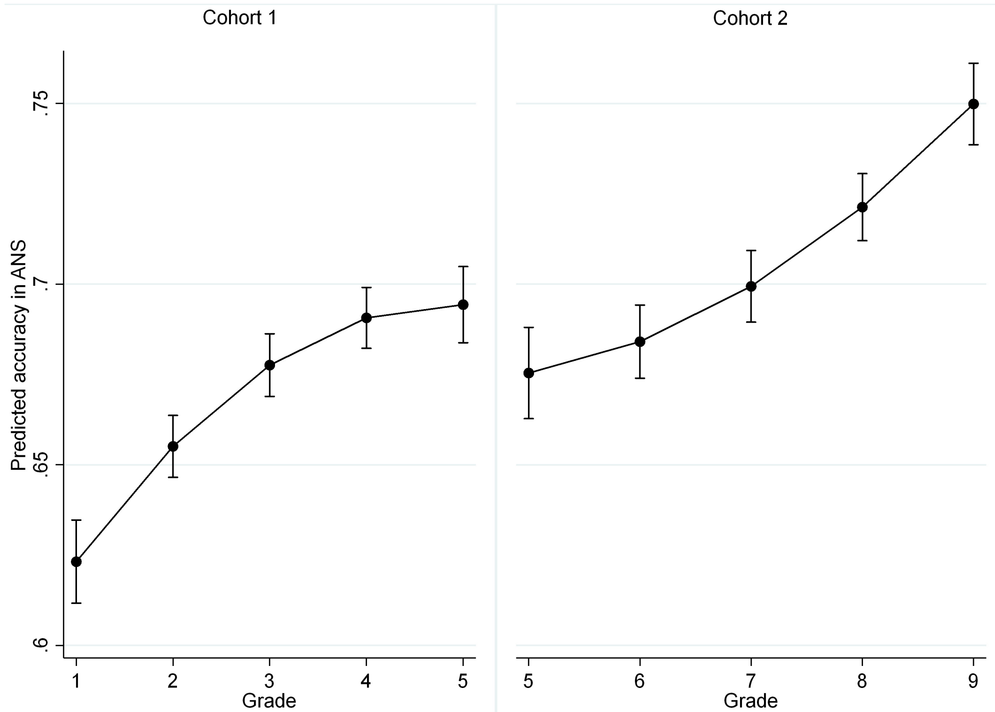


FIGURE 1 | Average predicted trajectories of changes in ANS accuracy for two cohorts (with 95% CI).

The results of the ME growth models of the RT in the non-symbolic comparison test in Cohort 2 (grade 5–grade 9) are presented in **Table 7**.

The analysis results revealed that the model with non-linear changes and a random slope for the variables "time" and "time^2" fit the data better than the other models. The RT decreased from grade 5 to grade 7 and then decreased more slowly thereafter. The post-estimation results revealed that there was no difference in RT between grades 7–9 (**Supplementary Table 2**). The covariance between the individual intercept and the slope of the variable "time" was negative, while the covariance between the intercept and slope of "time^2" was positive. This finding indicated that the individuals who had a larger RT in grade 5 demonstrated a larger decrease in RT from grade 5 to grade 6 and a larger deceleration later (**Supplementary Figure 4**).

The comparison of the average predicted trajectories of the RT in the non-symbolic comparison test and post-estimation revealed that the changes in RT across grades 5–9 were less prominent than those across grades 1–5 (**Figure 2**).

Notably, no significant differences in RT in the non-symbolic comparison test in grade 5 were observed between the two cohorts.

Changes in Accuracy and RT in Small and Large Ratio Bins

Considering the difference in accuracy between the small and large ratio bins, we inspected the growth trajectories in the easiest ratio (0.30–0.60) and hardest ratio (0.85–0.87). The results of Cohort 1 are presented in **Table 8**. The results revealed that in both ratio bins, accuracy increased according to a non-linear pattern. In the easiest ratio, the model with a random slope fit the data better than the model with a fixed slope. Hence, there was significant between-individual variability in the rate of change in accuracy. In the large ratio bin, the model with a random slope did not fit the data better than the model with a fixed slope. Therefore, the individual differences in the rate of change were not significant.

Although the pattern of the changes was the same in the two ratio bins, the coefficient of the variable "time" was higher in the easiest ratio, indicating a larger growth in accuracy in the easiest ratio. However, the absolute value of the negative regression coefficient of the "time squared" variable was smaller in the hardest ratio, indicating less slowing in growth (**Supplementary Figure 5**).

The results of the analysis of the changes in accuracy in Cohort 2 (grades 5–9) in two ratio bins separately are presented in **Table 9**. The analysis revealed that in the easiest ratio, accuracy increased according to a non-linear pattern. In particular, post-estimation revealed that accuracy did not increase from grade 5 to grade 7, but the difference between grade 7 and 8 became significant. The analysis also revealed that in the hardest ratio, accuracy increased according to a linear pattern. Hence, in the second cohort, the patterns of accuracy changes differed between the easiest and hardest ratio bins (**Supplementary Figure 6**).

The results of the analysis of the RT changes in the two ratio bins in Cohort 1 (grades 1–5) are presented in **Table 10**. The results of the analysis of the pattern of RT changes in

TABLE 6 | Cohort 1: results of ME growth models for changes in ANS RT (in sec.) from grade 1 to grade 5.

Variables	Baseline	Model 1	Model 2	Model 3
	Intercept-only	Linear growth	Non-linear growth	Model with random slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	1.22*** (0.02)	1.49*** (0.02)	1.52*** (0.02)	1.49*** (0.03)
Time		−0.13*** (0.01)	−0.17*** (0.02)	−0.13*** (0.01)
Time^2			0.01 (0.01)	
Random effect				
Intercept variance	0.04	0.05	0.05	0.15
Residuals	0.11	0.08	0.08	0.06
Slope variance (time)				0.01
Covariance between intercept and slope (time)				−0.03
Log-likelihood	−469.16	−300.47	−298.65	−245.37
LR test (Δ df)		337.38*** (1)	3.65 (1)	112.46*** (2)
ICC	0.26			

****p* < 0.001.*

the easiest ratio revealed that from grade 1 to grade 5, the RT decreased according to a linear pattern. The model with a random slope fit the data better; thus, there was significant between-individual variability in the rate of change in the RT in the easiest ratio. The analysis also demonstrated that the RT significantly decreased in the hardest ratio bin according to a linear pattern. In general, the patterns of change did not significantly differ between the easiest and hardest ratio bins in Cohort 1 (grades 1–5) (**Supplementary Figure 7**).

The results of the analysis of the changes in the RT of the correct answers in the two ratio bins in Cohort 2 (grades 5–9) are presented in **Table 11**. The results indicated that the RT in the easiest ratio decreased from grade 5 to grade 9 according to a non-linear pattern as follows: from grade 5 to grade 7, the RT significantly decreased, but these changes slowed thereafter. This pattern was also identified in the hardest ratio bin. In general, in Cohort 2, the patterns of changes in the RT did not significantly vary between the easiest and hardest ratio bins (**Supplementary Figure 8**). However, there was a tendency of increasing differences in RT between the two ratio bins.

How Do the Changes in ANS Accuracy and RT Relate to Each Other?

Next, we estimated the correlations between individual changes in ANS accuracy and RT. For each individual, the deviations from the average value of the time changes in accuracy and RT were calculated. Positive individual deviation values for accuracy indicated that the individual had a larger growth in ANS accuracy than the sample mean. Positive individual deviation values for ANS RT indicated that the individual had a slower decrease in ANS RT than the sample mean.

In Cohort 1 (grades 1–5), the correlation between the individual deviation in accuracy and RT was negative ($r = -0.18$, $p < 0.001$). This finding indicated that the individuals who demonstrated a faster decrease in RT had a greater

TABLE 7 | Cohort 2: results of ME growth models for changes in ANS RT (in sec.) from grade 5 to grade 9.

Variables	Baseline	Model 1	Model 2	Model 3	Model 3a
	Intercept-only	Linear growth	Non-linear growth	Model with random slope1	Model with random slope2
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect					
Constant	0.92*** (0.01)	1.00*** (0.01)	1.05*** (0.01)	1.05*** (0.02)	1.04*** (0.02)
Time		−0.04*** (0.004)	−0.14*** (0.01)	−0.14*** (0.02)	−0.14*** (0.02)
Time^2			0.02*** (0.003)	0.03*** (0.003)	0.02*** (0.003)
Random effect					
Intercept variance	0.02	0.02	0.02	0.04	0.04
Residuals	0.03	0.03	0.03	0.02	0.02
Slope variance (time)				0.002	0.02
Slope variance (time^2)					0.001
Covariance between intercept and slope (time)				−0.007	−0.02
Covariance between intercept and slope (time^2)					0.002
Covariance between slope (time) and slope (time^2)					−0.004
Log-likelihood	131.94	172.40	198.99	219.75	234.05
LR test (Δ df)		80.92*** (1)	53.19*** (1)	41.51*** (2)	28.59*** (3)
ICC	0.31				

****p* < 0.001.*

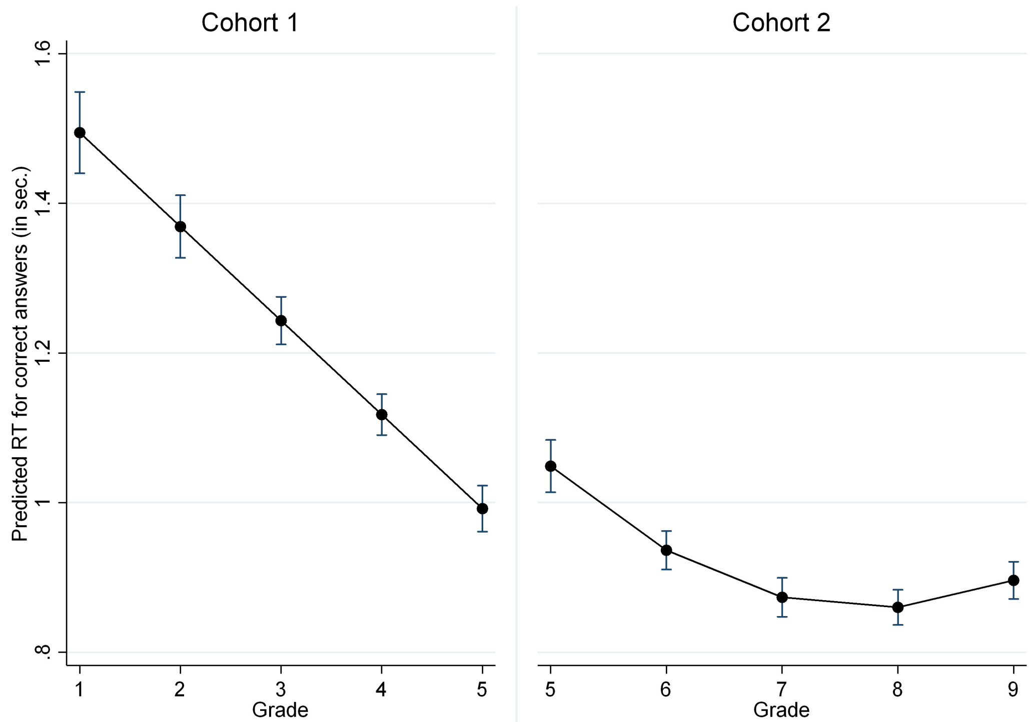


FIGURE 2 | Average predicted trajectories of changes in ANS RT for two cohorts (with 95% CI).

increase in accuracy, although the correlation was weak (**Supplementary Figure 9**).

In Cohort 2 (grades 5–9), there were significant individual differences in the slopes of the variables "time" and time^2" using the RT as the outcome; thus, two correlation coefficients were estimated. The individual deviations in the slope of the variable "time" in accuracy and RT were positively correlated ($r = 0.44$, $p < 0.001$), whereas the correlation between the individual slope of "time" in accuracy and the slope of "time ^{2}" in RT was negative ($r = -0.34, p < 0.001$). This finding indicated that the individuals who demonstrated a greater growth in accuracy had a smaller decrease in RT, but they exhibited less deceleration in the RT changes (**Supplementary Figure 10**).

Notably, the correlation between the changes in accuracy and RT in Cohort 1 (grades 1–5) was weaker than that in Cohort 2 (grades 5–9).

How Do the Changes in General PS Correlate With the Changes in ANS Accuracy and RT?

To estimate the extent to which the changes in general PS explain the changes in ANS accuracy and RT, we added general PS as a predictor of ANS accuracy and RT. The results are presented in **Table 12**. The analysis revealed that in Cohort 1, the changes in RT and accuracy were partially explained by the changes in general PS, although the changes in both accuracy and RT remained significant. A faster general PS was positively associated with higher accuracy and smaller RT in the non-symbolic comparison test. In Cohort 2, general PS was not correlated with RT in the ANS test and did not explain the changes in RT but was significantly correlated with accuracy.

Next, we estimated the developmental changes in general PS in Cohort 1 (**Supplementary Table 4**) and Cohort 2 (**Supplementary Table 5**). The analysis revealed that general PS increased from grade 1 to grade 5 according to a non-linear pattern and that there was significant between-individual variability in the rate of change. The results of Cohort 2 revealed that general PS improved from grade 5 to grade 9 according to a linear pattern.

Next, we compared the patterns of changes in non-symbolic and general PS (**Figure 3**). The analysis revealed that general PS changed in a non-linear pattern from grade 1 to grade 5, whereas non-symbolic PS changed in a linear pattern. In contrast, in Cohort 2, general PS changed linearly, whereas non-symbolic PS changed non-linearly. Notably, there was a significant difference in RT in the general PS test in grade 5 between Cohort 1 and Cohort 2.

DISCUSSION

This study aimed to estimate developmental changes in ANS precision from grade 1 to grade 5 and from grade 5 to grade 9 using longitudinal data from two cohorts of Russian children. Previously, investigations of the development of ANS precision were mostly based on evaluations of accuracy

TABLE 8 | Cohort 1: results of ME growth models for changes in ANS accuracy for the easiest (0.30–0.60) and hardest (0.85–0.87) ratio bins from grade 1 to grade 5.

Variables	Bin 1: 0.30–0.60		Bin 5: 0.85–0.87	
	Baseline	Model 3	Baseline	Model 2
	Intercept-only	Non-linear growth with random slope	Intercept-only	Non-linear growth with fixed slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	0.82*** (0.01)	0.74*** (0.01)	0.61*** (0.004)	0.58*** (0.01)
Time		0.05*** (0.01)		0.03*** (0.007)
Time^2		−0.007** (0.002)		−0.004* (0.002)
Random effect				
Intercept variance	0.0053	0.012	0.0015	0.0016
Residuals	0.0147	0.011	0.009	0.009
Slope variance (time)		0.00058		
Covariance between intercept and slope		−0.00197		
Log-likelihood	637.07	700.82	921.67	937.94
LR test (Δ df)		19.11*** (2) vs. model with fixed slope		3.86* (1) vs. model with linear growth
ICC	0.27		0.14	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

(e.g., Tikhomirova et al., 2019) or the Weber fraction, which is an indicator of ANS precision that is highly correlated with accuracy (e.g., Halberda et al., 2012; Inglis and Gilmore, 2014; Tosto et al., 2017). Less often, studies have used measures based on RT to estimate age-related differences in the ANS (Halberda et al., 2012). However, following the findings of previous studies using different measures of ANS precision (e.g., Dietrich et al., 2015, 2016), we assumed that an inspection of the developmental patterns of both accuracy and RT in non-symbolic comparison tests might provide important insight ANS development. Hence, we inspected the developmental patterns of ANS precision using two measures, i.e., the proportion of correct answers and mean RT of correct answers. To account for one of the main features of non-symbolic representations, i.e., ratio dependence, we also estimated the mean RT and accuracy in five ratio bins separately. We aimed to compare the developmental patterns of accuracy and RT between the smallest (easiest) ratio and the largest (hardest) ratio.

The analysis revealed that accuracy decreased as the ratio between the two compared sets increased, and in the largest ratio, accuracy was significantly lower than that in the smallest ratio. Furthermore, the difference in RT between the ratio bins was less impressive than that in accuracy. This finding indicated that the sensitivity to increasing ratios between the compared arrays manifested in decreasing accuracy, but the RT changed to a lesser extent.

The estimation of the developmental changes in accuracy in the two cohorts revealed that accuracy increased from grade 1 to grade 3 and from grade 7 to grade 9 but did not significantly change from grade 3 to grade 7. In both cohorts, the model with the quadratic patterns of changes fit the data better than the model with linear changes. The pattern of quadratic changes in Cohort 1 (grades 1–5) indicated faster growth in ANS accuracy and then slower changes. In Cohort 2 (grades 5–9), the opposite pattern was found as follows: the insignificant growth from grade 5 to grade 7 was replaced by growth in accuracy from grade 7 to grade 9. The analysis also revealed significant inter-individual changes in the rate of change in accuracy in both cohorts. Notably, the obtained quadratic pattern of changes fit better than the linear pattern in a restricted period only. The generalization of these patterns of changes to a wider period should be performed with caution. Quadratic models imply U-shaped trajectories in development, but this trajectory can manifest later in development. In this study, the quadratic pattern revealed that growth slowed (Cohort 1) or accelerated (Cohort 2).

The RT significantly decreased from grade 1 to grade 5 in a linear pattern. In the second cohort (grades 5–9), the changes

TABLE 9 | Cohort 2: results of ME growth models for changes in ANS accuracy for the easiest (0.30–0.60) and hardest (0.85–0.87) ratio bins from grade 5 to grade 9.

Variables	Bin 1: 0.30–0.60		Bin 5: 0.85–0.87	
	Baseline	Model 3	Baseline	Model 3
	Intercept-only	Non-linear growth with random slope	Intercept-only	Linear growth with random slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	0.86*** (0.01)	0.82*** (0.01)	0.65*** (0.004)	0.61*** (0.01)
Time		0.0015 (0.01)		0.015*** (0.002)
Time^2		0.006** (0.002)		
Random effect				
Intercept variance	0.005	0.010	0.002	0.005
Residuals	0.012	0.010	0.0085	0.0072
Slope variance (time)		0.0003		0.0002
Covariance between intercept and slope		−0.001		−0.0008
Log-likelihood	604.55	664.52	797.17	827.42
LR test (Δ df)		16.56*** (2) vs. model with non-linear growth and fixed slope		9.23** (2) vs. model with linear growth and fixed slope
ICC	0.31		0.22	

$^{***}p < 0.001$, $^{**}p < 0.01$.

TABLE 10 | Cohort 1: results of ME growth models for changes in ANS RT (in sec.) for the easiest (0.30–0.60) and hardest (0.85–0.87) ratio bins from grade 1 to grade 5.

Variables	Bin 1: 0.30–0.60		Bin 5: 0.85–0.87	
	Baseline	Model 3	Baseline	Model 3
	Intercept-only	Linear growth with random slope	Intercept-only	Linear growth with random slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	1.21*** (0.01)	1.50*** (0.02)	1.25*** (0.02)	1.51*** (0.03)
Time		−0.14*** (0.01)		−0.12*** (0.01)
Time^2				
Random effect				
Intercept variance	0.027	0.12	0.05	0.19
Residuals	0.12	0.06	0.14	0.08
Slope variance (time)		0.005		0.01
Covariance between intercept and slope		−0.02		−0.04
Log-likelihood	−468.83	−200.31	−605.33	−434.36
LR test (Δ df)		102.08*** (2) vs. model with linear growth and fixed slope		102.94*** (2) vs. model with linear growth and fixed slope
ICC	0.19		0.27	

***$p < 0.001$.

TABLE 11 | Cohort 2: results of ME growth models for changes in ANS RT (in sec.) for the easiest (0.30–0.60) and hardest (0.85–0.87) ratio bins from grade 5 to grade 9.

Variables	Bin 1: 0.30–0.60		Bin 5: 0.85–0.87	
	Baseline	Model 3	Baseline	Model 3
	Intercept-only	Non-linear growth with random slope	Intercept-only	Non-linear growth with random slope
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	0.87*** (0.01)	1.03*** (0.02)	0.94*** (0.01)	1.07*** (0.02)
Time		−0.15*** (0.01)		−0.15*** (0.02)
Time^2		0.023*** (0.003)		0.03*** (0.004)
Random effect				
Intercept variance	0.01	0.035	0.02	0.05
Residuals	0.036	0.021	0.05	0.03
Slope variance (time)		0.002		0.003
Covariance between intercept and slope		−0.007		−0.009
Log-likelihood	136.51	277.65	−26.88	34.16
LR test (Δ df)		63.31*** (2) vs. model with non-linear growth and fixed slope		37.89*** (2) vs. model with non-linear growth and fixed slope
ICC	0.22		0.30	

***$p < 0.001$.

in RT followed a non-linear pattern as follows: these changes occurred more rapidly from grade 5 to grade 7 and then slowed. Our study confirmed that ANS accuracy increased and RT decreased across development; thus, at the end of secondary school, the pupils demonstrated higher accuracy and shorter RT than the first-graders. These results are consistent with several studies demonstrating that adults have lower RTs and higher accuracy in ANS tests (e.g., Halberda et al., 2012). Furthermore, the period during which accuracy and RT change in one direction (increase in accuracy and decrease in RT) take turns with periods during which changes in RT may continue, while accuracy is stabilized, and vice versa.

The combination of changes in ANS accuracy and RT allows us to identify three stages of developmental changes in ANS precision across 9 years of formal schooling. The first stage (grade 1–grade 3, age from 7 to 9 years) was characterized by faster increases in accuracy and speed of non-symbolic comparison. During the second stage (grade 3–grade 7, age from 9–13), accuracy stabilized, while the speed of non-symbolic comparisons continued to increase. During the third stage (grade 7–grade 9, 13–15 years), ANS accuracy started to increase again, while ANS RT did not significantly change.

These findings indicate that at different developmental stages, changes in the precision of ANS manifest in different ANS measures, which should be considered. It is possible that at the beginning of formal education, changes in the precision of ANS manifest in both accuracy and RT, but later, growing precision mostly manifests in decreased RT but not increased accuracy. At the end of secondary school (grades 7–9, age range 13–15 years), in turn, changes in RT might not reflect changes in ANS precision, whereas growth in accuracy might indicate growth in ANS precision during this stage of development.

Although we did not directly estimate the NRE and its changes, we can compare the developmental trajectories between the easiest and hardest ratio bins. The inspection of the changes in accuracy in the two ratio bins revealed that from grade 1 to grade 5, the changes in the easiest ratio bin were larger than those in the hardest ratio bin, although in both ratio bins, non-linear patterns of changes were identified. In Cohort 2 (grades 5–9), the patterns of the change in accuracy differed between the two ratio bins. The easiest ratio changes followed a non-linear pattern with acceleration of growth, while accuracy in the hardest ratio bin changed linearly. Notably, in both cohorts, the changes in accuracy were more prominent in the easiest ratio. This finding might indicate that the increased accuracy in the non-symbolic comparison test was driven by an increase in accuracy in easier tasks. These results are likely to indicate a slight increase in the NRE as this increase occurs on account of growth in accuracy in trials with the easy ratio.

TABLE 12 | Results of ME growth models for changes in ANS accuracy and RT with general PS (in sec.) as a predictor.

Variables	Cohort 1 (grades 1–5)		Cohort 2 (grades 5–9)	
	Accuracy	RT	Accuracy	RT
	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Fixed effect				
Constant	0.63*** (0.01)	1.48*** (0.03)	0.68*** (0.01)	1.07*** (0.02)
Time	0.03*** (0.005)	−0.12*** (0.01)	−0.001 (0.005)	−0.14*** (0.02)
Time^2	−0.004** (0.001)		0.004*** (0.001)	0.02*** (0.003)
General PS (in Z-scores)	−0.007* (0.003)	0.025* (0.01)	−0.02*** (0.003)	−0.02 (0.05)
Random effect				
Intercept variance	0.004	0.15	0.005	0.04
Residuals	0.004	0.06	0.003	0.02
Slope variance (time)	0.0002	0.01	0.0002	0.02
Slope variance (time^2)				0.001
Covariance between intercept and slope (time)	−0.0005	−0.03	−0.0006	−0.02
Covariance between intercept and slope (time^2)				0.002
Covariance between slope (time) and slope (time^2)				−0.004
Log-likelihood	1273.99	−242.72	1111.80	234.17
LR test (Δ df)	6.38* (1) (vs. Model 3)	5.31* (1) (vs. Model 3)	45.80*** (1) (vs. Model 3)	0.24 (1) (vs. Model 3a)

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

There are controversial findings regarding the development of the NRE. Some studies have demonstrated that the NRE is reduced across age (Holloway and Ansari, 2009), while other studies have demonstrated increases in the NRE (Lyons et al., 2015). Several studies also found that the NRE or NDE were stable across time (Reynvoet et al., 2009; Defever et al., 2011). The differences in the obtained findings might be related to different formats of magnitude (symbolic or non-symbolic), different types of tasks (priming vs. comparison) or different formats of stimulus presentation (paired vs. intermixed format) in the non-symbolic comparison task. Particularly, it has been demonstrated that the NRE under paired conditions was stronger than that under intermixed conditions (Price et al., 2012). It has also been demonstrated that the distance effect in priming tasks was stable across age (Defever et al., 2011), while the distance effect in comparison tasks decreased (Holloway and Ansari, 2008). In general, it might be concluded that the NRE is sensitive to the format of tasks and cannot be considered a reliable measure of ANS precision and its development.

In addition, the results of the current study indicated that accuracy and RT had different levels of inter-individual variability. The ICC value of accuracy was higher than that of RT in both cohorts (for accuracy, the ICC value was 0.36 and 0.39 in Cohort 1 and Cohort 2, respectively, whereas for RT, the ICC value was 0.26 and 0.31). This finding indicates that individuals exhibited variations in accuracy in the ANS test to a greater extent than they exhibited variations in RT and that RT was a less stable measure of ANS precision than accuracy.

The different roles of accuracy and RT were considered within the diffusion model (Ratcliff, 2002; Park and Starns, 2015; Ratcliff et al., 2016). The diffusion model considers each task a decision process that can be performed based on the noisy accumulation of information. Several components of decision processes were identified, including the drift rate (the rate of the accumulation of information available for use in a decision), boundary settings (boundary of correct or incorrect responses) and non-decision processes. Ratcliff et al. (2015) demonstrated that in numerical tasks, accuracy is largely determined by the drift rate, whereas the RT is determined by boundary settings. It was also shown that the slower RT of children than that of young adults could be explained by wider boundary separation and non-decision processes. For example, the reduction in the RT of older children compared to that of first-graders might be related to a decrease in the amount of time devoted to non-decision processes, such as stimulus encoding and response execution (Ratcliff et al., 2012).

It is possible to assume that the changes in accuracy and RT can be explained by different factors. The faster growth in ANS accuracy at the start of formal schooling might be associated with the acquisition of symbolic math skills and math knowledge, which may facilitate ANS development. Evidence suggests that education has a significant effect on ANS precision and that symbolic representation predicts the precision of non-symbolic representation (e.g., Piazza et al., 2013; Mussolin et al., 2014; Shusterman et al., 2016). In addition, pupils start to receive regular feedback from their teachers and parents during grades 1–2. Previous studies have demonstrated that feedback may improve ANS precision (e.g., DeWind and Brannon, 2012). Thus, children have the opportunity to adjust the system of non-symbolic representation at the start of formal schooling during the acquisition of symbolic math skills, and receiving feedback contributes to improvements in ANS precision.

The improvement in non-symbolic comparison might reflect the progressive automatization of access to non-symbolic representation. Ample evidence highlights the involvement of the IPS in the processing of numerosity in both symbolic and non-symbolic formats (e.g., Hubbard et al., 2008; Holloway and Ansari, 2010). It has been demonstrated that the involvement of the IPS in processing symbolic and non-symbolic numerosity increases across age (Ansari et al., 2005; Ansari and Dhital, 2006; Hubbard et al., 2008), while the activation of frontal areas decreases (e.g., Gullick and Wolford, 2013). Many studies have demonstrated a frontoparietal shift in numerical cognition, which likely reflects less recruitment of frontal areas associated with attention, working memory, and executive functions (Ansari et al., 2005; Rivera et al., 2005). Evidence indicates that slower individuals may require more prefrontal executive control than faster individuals to perform successfully (Rypma et al., 2006). Therefore, an increase in non-symbolic PS might

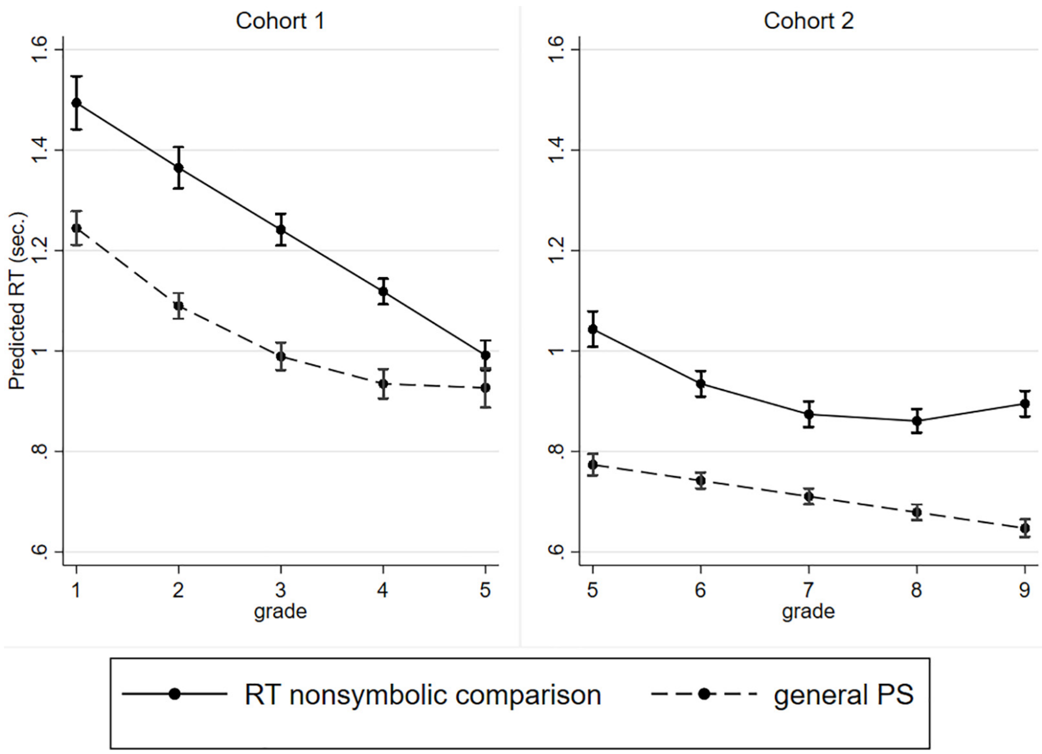


FIGURE 3 | Developmental changes in RT for the ANS test and for the general PS test (in second) with 95% CI.

reflect a reduced involvement of frontal areas during non-symbolic comparisons.

The difference in the mechanisms supporting changes in accuracy and RT was demonstrated in several studies of non-numerical processing. In particular, Santee and Egeth (1982) postulated that accuracy and RT reflect different perceptual processes in letter recognition tasks. Accuracy is more sensitive to the early perceptual stage of processing, whereas the RT is more sensitive to later perceptual processing. This difference was also confirmed in studies involving other perceptual and attentional tasks. For example, in somatosensory discrimination tasks, attentional cues have been found to affect accuracy and RT via different cognitive and neural processing methods (van Ede et al., 2012). The cueing effect on accuracy was explained by a preparatory process (increasing activity in the somatosensory cortex) only, whereas the effect of RT was additionally explained by a post-target process. Perri et al. (2014) conducted an EEG study involving an execution go/no-go task and demonstrated that speed and accuracy are processed by two interacting but separate neurocognitive systems. The authors identified groups of individuals according to their tendency to prefer speed or accuracy and considered event-related potential (ERP) components after a stimulus to highlight the different levels of perceptual processing-supported speed or accuracy tendency. It was demonstrated that baseline activity (before the stimulus appearance) in the supplementary motor area differentiates "speedy" and "slow" individuals, whereas activation of the right prefrontal cortex differentiates "accurate" and "inaccurate" groups. The analysis of post-stimulus activity revealed a difference in the P1 ERP component between the faster and slow groups and a difference in the N1 ERP component between the accurate and inaccurate groups. Considering the aforementioned studies, it is possible that differences in developmental changes in accuracy and RT in non-symbolic comparisons to some extent reflect differences in the maturation and development of two distinct neurocognitive systems. This suggestion can be verified in future longitudinal and neurophysiological studies.

In general, our findings confirm the results of previous studies demonstrating that RT-based measures do not reflect ANS precision in the same way as accuracy-based measures (Dietrich et al., 2016). Although the RT decreased over time, the interpretation of a faster RT as an indicator of a more precise ANS needs to be clarified. The present analysis revealed that in Cohort 1, the improvements in accuracy and speed were positively correlated; thus, the pupils who demonstrated higher growth in accuracy also demonstrated a higher rate of change in the RT. In the second cohort, the opposite pattern was revealed. The pupils who had a greater increase in accuracy demonstrated a lower rate of change in the RT. This finding might indicate that although a lower RT corresponded to older participants from a developmental perspective, it does not always reflect increased accuracy in non-symbolic representation.

This study also revealed that general PS and speed in non-symbolic comparison tasks increased across age. The improvement in both general and non-symbolic PS might be explained by the processes of neuronal axon myelination and synaptic pruning (the process of synapse elimination) (Travis, 1998; Chechik et al., 1999). The myelination of neurons results in more rapid neural computation through faster propagation of action potentials (Mabbott et al., 2006; Fields, 2008; Chevalier et al., 2015). It has also been shown that individual differences in general PS might be associated with regional connectivity, implying a central role of axonal structures in inter-individual activation differences (Rypma et al., 2006). Synaptic pruning leads to a reduction in unused pathways and the strengthening of used pathways (e.g., Chechik et al., 1998). It has been postulated that the process of pruning is driven by individual experience and allows an individual to respond faster to the unique environment in which s/he grows (e.g., Tierney and Nelson, 2009).

However, general PS and non-symbolic PS develop at different rates in different patterns. In Cohort 1, linear changes in non-symbolic RT and non-linear changes in general PS were identified. In Cohort 2, the opposite patterns were observed as follows: general PS developed linearly, while the RT in the non-symbolic comparison changed non-linearly. Moreover, in both cohorts, the changes in general PS did not eliminate the time changes in non-symbolic comparison RT. In addition, general PS was not associated with RT in the non-symbolic comparison in the pupils in grade 5 to grade 9. These findings might confirm the local trend hypothesis of PS development.

It is possible that the development of general PS forms the basis for the development of non-symbolic PS. For example, it has been shown that training in PS led to improvements in other cognitive functions (Takeuchi and Kawashima, 2012). The patterns of change in non-symbolic PS repeated the developmental patterns of general PS at a previous age. However, the opposite relationships might also exist, i.e., the development of general PS might combine the development of specific processes. To verify this suggestion, it is necessary to include more time points in longitudinal data and additional different tasks for the estimation of PS in different processes.

Notably, in this study, general PS was more correlated with accuracy than RT in the non-symbolic comparison test. On the one hand, these results might reflect the close relationships between general PS and other cognitive constructs measured by accuracy. For example, many studies have demonstrated that general PS is associated with intelligence and working memory (Fry and Hale, 1996; Sheppard and Vernon, 2008). Moreover, it has been shown that general PS is substantially correlated with untimed tests (Wilhelm and Schulze, 2002). It is possible that the association between accuracy in the non-symbolic comparison test and general PS is not explained by time restriction during the execution of a non-symbolic comparison test.

On the other hand, the association between accuracy in a non-symbolic comparison test and general PS can be explained by the specificity of the general PS test, which was considered in the current study. In the test used in the present study, the children were asked to press a key corresponding to a digit (1, 2, 3, or 4) appearing on the screen as fast and accurately as possible. The mean RT of the correct answers was used as an indicator of general PS. Therefore, symbolic math skills were utilized to some extent to execute this test. The link between the results of the RT test and the accuracy of the non-symbolic comparison test might be partially explained by their association with symbolic math skills.

The current study had some limitations regarding the test used for the estimation of ANS. Some authors suggest that in tasks involving non-symbolic comparison, individuals are affected by the visual properties of the arrays. Arrays of objects can be compared based on comparisons of visual properties, such as cumulative area or convex hull (Gebuis and Reynvoet, 2012; Gebuis et al., 2016). To confirm the effect of visual properties on accuracy in comparisons of two sets of dots, researchers have manipulated different visual properties and identified two types of trials. The first type was congruent trials in which the visual properties were positively correlated with the magnitude. The second type was incongruent trials in which the magnitude was negatively correlated with the visual properties (e.g., Gebuis and Reynvoet, 2012; Clayton et al., 2015; Gilmore et al., 2016). It was demonstrated that accuracy in such comparisons was higher and the RT was faster in congruent trials than incongruent trials (congruency effect) (e.g., Gebuis and Reynvoet, 2012; Szucs et al., 2013). The congruency effect was used to confirm that numerosity judgments are based on the estimation of the visual properties of stimuli (Gebuis and Reynvoet, 2012).

In the current version of the ANS test, all trials in the test were congruent, and the array that contained more dots had a larger cumulative area. Hence, this version of the ANS test can measure accuracy in both non-symbolic representation and estimation of visual cues. It has been shown that activation of brain areas involved in numerical processing does not significantly differ between congruent and incongruent trials (Wilkey et al., 2017). This finding might indicate that even in congruent trials, individual can estimate numerosity in parallel with visual cues. Moreover, we used a "blue-yellow dots" test with an intermixed format, and it has been demonstrated that the reliability of this test in the intermixed format is higher than that in the paired or sequential formats (Price et al., 2012). Based on previous findings, we propose that the obtained results reflect the developmental trends in non-symbolic comparisons to a large extent.

We also used the same version of the test each year. This approach has some advantages, such as the ability to directly compare accuracy and RT across years. The period between testing was relatively long (nearly 1 year), and feedback was not provided; thus, we can avoid the effect of memory or training on the results of the test.

CONCLUSION

This study is the first to estimate the longitudinal development of ANS precision based on an inspection of changes in both accuracy and RT. Our findings revealed that the developmental patterns of changes in ANS accuracy and RT were not synchronous, but an inspection of both measures might provide new insight into ANS development.

In general, three stages of ANS development were identified. During stage 1 (grade 1–grade 3, age 7–9 years), development was characterized by faster growth in accuracy and non-symbolic PS.

Stage 2 (grade 3–grade 7, age 9–13 years) was characterized by stability in accuracy and continuing increases in non-symbolic PS. During stage 3 (grade 7–grade 9, age 13–15), the opposite trend was revealed, i.e., accuracy started to increase, while PS stabilized. A speed-accuracy trade-off was identified at all time points. In general, the results of this study suggest that for a more informative investigation of ANS development, an inspection of both accuracy and RT is needed.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Psychological Institute of the Russian Academy of Education. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SM directs and received funding for the "Cross-cultural Longitudinal Analysis of Student Success" (CLASS) project. SM and TT conceived and designed the present study. YK and TT conducted the analyses and interpreted the results under the supervision of SM. YK drafted the manuscript. All authors discussed the results and implications and provided comments regarding the manuscript at all stages. All authors approved the final version of the manuscript for submission.

FUNDING

This study was supported by a grant (No. 17-78-30028) from the Russian Science Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.589305/full#supplementary-material

REFERENCES

Agrillo, C., Dadda, M., Serena, G., and Bisazza, A. (2009). Use of number by fish. *PLoS One* 4:e4786. doi: 10.1371/journal.pone.0004786

Anobile, G., Castaldi, E., Turi, M., Tinelli, F., and Burr, D. C. (2016). Numerosity but not texture-density discrimination correlates with math ability in children. *Dev. Psychol.* 52, 1206–1216. doi: 10.1037/dev0000155

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

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

Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28. doi: 10.21500/20112084.807

Bisanz, J., Danner, F., and Resnick, L. B. (1979). Changes with age in measures of processing efficiency. *Child Dev.* 50, 132–141. doi: 10.2307/1129049

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

Burr, D. C., Turi, M., and Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *J. Vis.* 10, 20–20. doi: 10.1167/10.6.20

Cantlon, J. F., and Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biol.* 5:e328. doi: 10.1371/journal.pbio.0050328

Chechik, G., Meilijson, I., and Ruppin, E. (1998). Synaptic pruning in development: a computational account. *Neural Comput.* 10, 1759–1777. doi: 10.1162/089976698300017124

Chechik, G., Meilijson, I., and Ruppin, E. (1999). Neuronal regulation: a mechanism for synaptic pruning during brain maturation. *Neural Comput.* 11, 2061–2080. doi: 10.1162/089976699300016089

Chevalier, N., Kurth, S., Doucette, M. R., Wiseheart, M., Deoni, S. C., Dean, D. C. III, et al. (2015). Myelination is associated with processing speed in early childhood: preliminary insights. *PLoS One* 10:e0139897. doi: 10.1371/journal.pone.0139897

Chopra, S., Shaw, M., Shaw, T., Sachdev, P. S., Anstey, K. J., and Cherbuin, N. (2018). More highly myelinated white matter tracts are associated with faster processing speed in healthy adults. *Neuroimage* 171, 332–340. doi: 10.1016/j.neuroimage.2017.12.069

Clayton, S., Gilmore, C., and Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychol.* 161, 177–184. doi: 10.1016/j.actpsy.2015.09.007

Coertjens, L., Donche, V., De Maeyer, S., Vanthournout, G., and Van Petegem, P. (2017). To what degree does the missing-data technique influence the estimated growth in learning strategies over time? A tutorial example of sensitivity analysis for longitudinal data. *PLoS One* 12:e0182615. doi: 10.1371/journal.pone.0182615

Coyle, T. R., Pillow, D. R., Snyder, A. C., and Kochunov, P. (2011). Processing speed mediates the development of general intelligence (g) in adolescence. *Psychol. Sci.* 22, 1265–1269. doi: 10.1177/0956797611418243

Curran, P. J., Obeidat, K., and Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* 11, 121–136. doi: 10.1080/15248371003699969

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

Deary, I. J., Der, G., and Ford, G. (2001). Reaction times and intelligence differences: a population-based cohort study. *Intelligence* 29, 389–399. doi: 10.1016/S0160-2896(01)00062-9

Defever, E., Sasanguie, D., Gebuis, T., and Reynvoet, B. (2011). Children's representation of symbolic and nonsymbolic magnitude examined with the priming paradigm. *J. Exp. Child Psychol.* 109, 174–186. doi: 10.1016/j.jecp.2011.01.002

Dehaene, S. (2001). Précis of the number sense. *Mind Lang.* 16, 16–36.

DeWind, N. K., and Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Front. Hum. Neurosci.* 6:68. doi: 10.3389/fnhum.2012.00068

Dietrich, J. F., Huber, S., Klein, E., Willmes, K., Pixner, S., and Moeller, K. (2016). A systematic investigation of accuracy and response time based measures used to index ANS acuity. *PLoS One* 11:e0163076. doi: 10.1371/journal.pone.0163076

Dietrich, J. F., Huber, S., and Nuerk, H. C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS)–a research review. *Front. Psychol.* 6:295. doi: 10.3389/fpsyg.2015.00295

Dodonova, Y. A., and Dodonov, Y. S. (2013). Faster on easy items, more accurate on difficult ones: cognitive ability and performance on a task of varying difficulty. *Intelligence* 41, 1–10. doi: 10.1016/j.intell.2012.10.003

Duncan, T. E., and Duncan, S. C. (2009). The ABC's of LGM: an introductory guide to latent variable growth curve modeling. *Soc. Pers. Psychol. Compass* 3, 979–991. doi: 10.1111/j.1751-9004.2009.00224.x

Fields, R. D. (2008). White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* 31, 361–370. doi: 10.1016/j.tins.2008.04.001

Fry, A. F., and Hale, S. (1996). Processing speed, working memory, and fluid intelligence: evidence for a developmental cascade. *Psychol. Sci.* 7, 237–241. doi: 10.1111/j.1467-9280.1996.tb00366.x

Gebuis, T., Kadosh, R. C., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. *Acta Psychol.* 171, 17–35. doi: 10.1016/j.actpsy.2016.09.003

Gebuis, T., and Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PLoS One* 7:e37426. doi: 10.1371/journal.pone.0037426

Gilmore, C., Cragg, L., Hogan, G., and Inglis, M. (2016). Congruency effects in dot comparison tasks: convex hull is more important than dot area. *J. Cogn. Psychol.* 28, 923–931. doi: 10.1080/20445911.2016.1221828

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

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

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

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

Hale, S. (1990). A global developmental trend in cognitive processing speed. *Child Dev.* 61, 653–663. doi: 10.1111/j.1467-8624.1990.tb02809.x

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

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

Holloway, I. D., and Ansari, D. (2010). Developmental specialization in the right intraparietal sulcus for the abstract representation of numerical magnitude. *J. Cogn. Neurosci.* 22, 2627–2637. doi: 10.1162/jocn.2009.21399

Hubbard, E. M., Diester, I., Cantlon, J. F., Ansari, D., Van Opstal, F., and Troiani, V. (2008). The evolution of numerical cognition: from number neurons to linguistic quantifiers. *J. Neurosci.* 28, 11819–11824. doi: 10.1523/JNEUROSCI.3808-08.2008

Inglis, M., and Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychol.* 145, 147–155. doi: 10.1016/j.actpsy.2013.11.009

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Amsterdam: Elsevier.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychol. Bull.* 109, 490–501. doi: 10.1037/0033-2909.109.3.490

Kail, R. (2000). Speed of information processing: developmental change and links to intelligence. *J. Sch. Psychol.* 38, 51–61. doi: 10.1016/S0022-4405(99)00036-9

Kail, R. V., and Ferrer, E. (2007). Processing speed in childhood and adolescence: longitudinal models for examining developmental change. *Child Dev.* 78, 1760–1770. doi: 10.1111/j.1467-8624.2007.01088.x

Kail, R. V., Lervåg, A., and Hulme, C. (2016). Longitudinal evidence linking processing speed to the development of reasoning. *Dev. Sci.* 19, 1067–1074. doi: 10.1111/desc.12352

Kail, R. V., and Miller, C. A. (2006). Developmental change in processing speed: domain specificity and stability during childhood and adolescence. *J. Cogn. Dev.* 7, 119–137. doi: 10.1207/s15327647jcd0701_6

Kuzmina, Y., Tikhomirova, T., Lysenkova, I., and Malykh, S. (2020). Domain-general cognitive functions fully explained growth in nonsymbolic magnitude representation but not in symbolic representation in elementary school children. *PLoS One* 15:e0228960. doi: 10.1371/journal.pone.0228960

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. Diff.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *J. Am. Stat. Assoc.* 83, 1198–1202. doi: 10.1080/01621459.1988.10478722

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *J. Neurolinguist.* 24, 583–591. doi: 10.1016/j.jneuroling.2011.02.004

Lyons, I. M., Nuerk, H. C., and Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *J. Exp. Psychol. Gen.* 144:1021. doi: 10.1037/xge0000094

Mabbott, D. J., Noseworthy, M., Bouffet, E., Laughlin, S., and Rockel, C. (2006). White matter growth as a mechanism of cognitive development in children. *Neuroimage* 33, 936–946. doi: 10.1016/j.neuroimage.2006.07.024

Mussolin, C., Mejias, S., and Noël, M. P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition* 115, 10–25. doi: 10.1016/j.cognition.2009.10.006

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

Nettelbeck, T., and Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Pers. Individ. Diff.* 48, 379–384. doi: 10.1016/j.paid.2009.10.032

Neubauer, A. C. (1990). Speed of information processing in the Hick paradigm and response latencies in a psychometric intelligence test. *Pers. Individ. Diff.* 11, 147–152. doi: 10.1016/0191-8869(90)90007-e

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

Norris, J. E., and Castronovo, J. (2016). Dot display affects approximate number system acuity and relationships with mathematical achievement and inhibitory control. *PLoS One* 11:e0155543. doi: 10.1371/journal.pone.0155543

Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Dev. Sci.* 21:e12533. doi: 10.1111/desc.12533

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. doi: 10.1037/a0029472

Olivers, C. N., and Watson, D. G. (2008). Subitizing requires attention. *Vis. Cogn.* 16, 439–462. doi: 10.1080/13506280701825861

Park, J., and Starns, J. J. (2015). The approximate number system acuity redefined: a diffusion model approach. *Front. Psychol.* 6:1955. doi: 10.3389/fpsyg.2015.01955

Perri, R. L., Berchicci, M., Spinelli, D., and Di Russo, F. (2014). Individual differences in response speed and accuracy are associated to specific brain activities of two interacting systems. *Front. Behav. Neurosci.* 8:251. doi: 10.3389/fnbeh.2014.00251

Pezzuti, L., Lauriola, M., Borella, E., De Beni, R., and Cornoldi, C. (2019). Working memory and processing speed mediate the effect of age on a general ability construct: evidence from the Italian WAIS-IV standardization sample. *Pers. Individ. Diff.* 138, 298–304. doi: 10.1016/j.paid.2018.10.016

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

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

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

Pomè, A., Anobile, G., Cicchini, G. M., and Burr, D. C. (2019). Different reaction-times for subitizing, estimation, and texture. *J. Vis.* 19, 14–14. doi: 10.1167/19.6.14

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

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: fitting real data and failing to fit fake but plausible data. *Psychon. Bull. Rev.* 9, 278–291. doi: 10.3758/bf03196283

Ratcliff, R., Love, J., Thompson, C. A., and Opfer, J. E. (2012). Children are not like older adults: a diffusion model analysis of developmental changes in speeded responses. *Child Dev.* 83, 367–381. doi: 10.1111/j.1467-8624.2011.01683.x

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends Cogn. Sci.* 20, 260–281. doi: 10.1016/j.tics.2016.01.007

Ratcliff, R., Thompson, C. A., and McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition* 137, 115–136. doi: 10.1016/j.cognition.2014.12.004

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., and Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychol. Sci.* 19, 607–614. doi: 10.1111/j.1467-9280.2008.02130.x

Reynvoet, B., De Smedt, B., and Van den Bussche, E. (2009). Children's representation of symbolic magnitude: the development of the priming distance effect. *J. Exp. Child Psychol.* 103, 480–489. doi: 10.1016/j.jecp.2009.01.007

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

Rypma, B., Berger, J. S., Prabhakaran, V., Bly, B. M., Kimberg, D. Y., Biswal, B. B., et al. (2006). Neural correlates of cognitive efficiency. *Neuroimage* 33, 969–979. doi: 10.1016/j.neuroimage.2006.05.065

Santee, J. L., and Egeth, H. E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? *J. Exp. Psychol. Hum. Percept. Perform.* 8:489. doi: 10.1037/0096-1523.8.4.489

Sasanguie, D., Defever, E., Van den Bussche, E., and Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychol.* 136, 73–80. doi: 10.1016/j.actpsy.2010.10.004

Scantlebury, N., Cunningham, T., Dockstader, C., Laughlin, S., Gaetz, W., Rockel, C., et al. (2014). Relations between white matter maturation and reaction time in childhood. *J. Int. Neuropsychol. Soc. JINS* 20, 99–112. doi: 10.1017/S1355617713001148

Schramm, P., and Rouder, J. (2019). Are reaction time transformations really beneficial?. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/9ksa6

Sheppard, L. D., and Vernon, P. A. (2008). Intelligence and speed of information-processing: a review of 50 years of research. *Pers. Individ. Diff.* 44, 535–551. doi: 10.1016/j.paid.2007.09.015

Shusterman, A., Slusser, E., Halberda, J., and Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS One* 11:e0153072. doi: 10.1371/journal.pone.0153072

Smets, K., Moors, P., and Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: implications for number processing? *Front. Psychol.* 7:66. doi: 10.3389/fpsyg.2016.00066

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

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

Takeuchi, H., and Kawashima, R. (2012). Effects of processing speed training on cognitive functions and neural systems. *Rev. Neurosci.* 23, 289–301. doi: 10.1515/revneuro-2012-0035

Tierney, A. L., and Nelson, C. A. III (2009). Brain development and the role of experience in the early years. *Zero Three* 30:9.

Tikhomirova, T., Kuzmina, Y., Lysenkova, I., and Malykh, S. (2019). Development of approximate number sense across the elementary school years: a cross-cultural longitudinal study. *Dev. Sci.* 22:e12823. doi: 10.1111/desc.12823

Tosto, M. G., Petrill, S. A., Malykh, S., Malki, K., Haworth, C., Mazzocco, M. M., et al. (2017). Number sense and mathematics: which, when and how? *Dev. Psychol.* 53:1924. doi: 10.1037/dev0000331

Travis, F. (1998). Cortical and cognitive development in 4th, 8th and 12th grade students: the contribution of speed of processing and executive functioning to cognitive development. *Biol. Psychol.* 48, 37–56. doi: 10.1016/s0301-0511(98)00005-2

van Ede, F., de Lange, F. P., and Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *J. Neurosci.* 32, 10408–10412. doi: 10.1523/JNEUROSCI.1337-12.2012

Vanbinst, K., Ghesquiere, 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.2012.01148.x

Vetter, P., Butterworth, B., and Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: evidence against a pre-attentive subitizing mechanism. *PLoS One* 3:e3269. doi: 10.1371/journal.pone.0003269

Viswanathan, P., and Nieder, A. (2013). Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11187–11192. doi: 10.1073/pnas.1308141110

Whelan, R. (2008). Effective analysis of reaction time data. *Psychol. Rec.* 58, 475–482. doi: 10.1007/BF03395630

Wilhelm, O., and Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. *Intelligence* 30, 537–554. doi: 10.1016/s0160-2896(02)00086-7

Wilkey, E. D., Barone, J. C., Mazzocco, M. M., Vogel, S. E., and Price, G. R. (2017). The effect of visual parameters on neural activation during nonsymbolic number comparison and its relation to math competency. *NeuroImage* 159, 430–442. doi: 10.1016/j.neuroimage.2017.08.023

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fpsyg.2021.584183

Numerical Magnitude Processing in Deaf Adolescents and Its Contribution to Arithmetical Ability

Lilan Chen[1]*, Yan Wang[2] and Hongbo Wen[3]

[1] School of Psychology, Hainan Normal University, Haikou, China, [2] Faculty of Education, Beijing Normal University, Beijing, China, [3] Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, Beijing, China

Although most deaf individuals could use sign language or sign/spoken language mix, hearing loss would still affect their language acquisition. Compensatory plasticity holds that the lack of auditory stimulation experienced by deaf individuals, such as congenital deafness, can be met by enhancements in visual cognition. And the studies of hearing individuals have showed that visual form perception is the cognitive mechanism that could explain the association between numerical magnitude processing and arithmetic computation. Therefore, we examined numerical magnitude processing and its contribution to arithmetical ability in deaf adolescents, and explored the differences between the congenital and acquired deafness. 112 deaf adolescents (58 congenital deafness) and 58 hearing adolescents performed a series of cognitive and mathematical tests, and it was found there was no significant differences between the congenital group and the hearing group, but congenital group outperformed acquired group in numerical magnitude processing (reaction time) and arithmetic computation. It was also found there was a close association between numerical magnitude processing and arithmetic computation in all deaf adolescents, and after controlling for the demographic variables (age, gender, onset of hearing loss) and general cognitive abilities (non-verbal IQ, processing speed, reading comprehension), numerical magnitude processing could predict arithmetic computation in all deaf adolescents but not in congenital group. The role of numerical magnitude processing (symbolic and non-symbolic) in deaf adolescents' mathematical performance should be paid attention in the training of arithmetical ability.

Keywords: numerical magnitude representation, arithmetic computation, congenital deafness, acquired deafness, mathematical cognition

OPEN ACCESS

Edited by:
Fanli Jia,
Seton Hall University, United States

Reviewed by:
Andrej Kral,
Hannover Medical School, Germany
Emmanouela Terlektsi,
University of Birmingham, United Kingdom

****Correspondence:***
Lilan Chen
chenll22@163.com

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

Received: *17 September 2020*
Accepted: *02 March 2021*
Published: *25 March 2021*

Citation:
Chen L, Wang Y and Wen H (2021) Numerical Magnitude Processing in Deaf Adolescents and Its Contribution to Arithmetical Ability. Front. Psychol. 12:584183. doi: 10.3389/fpsyg.2021.584183

INTRODUCTION

Mathematical knowledge and ability play an important role in the successes of our social life (Ritchie and Bates, 2013), but most deaf individuals have some difficulty in acquisition of arithmetical skills even if they have the approximately same level of non-verbal intelligence as hearing peers (Braden, 1994; Moreno, 2000). Many studies have shown the close association between numerical magnitude processing and mathematical ability

(LeFevre et al., 2010; De Smedt et al., 2013; Sasanguie et al., 2013; Fazio et al., 2014; Linsen et al., 2015), although few studies on the arithmetical abilities of deaf individuals (Masataka, 2006; Andin et al., 2014, 2020). It has been found in many studies that the poorer performance of deaf individuals in mathematics has generally been associated with their reduced language abilities (Kelly and Gaustad, 2007; Wu et al., 2013; Huber et al., 2014; Vitova et al., 2014).

Although most deaf individuals could use sign language or sign/spoken language mix, hearing loss would still affect their language acquisition (Kennedy et al., 2006; Elizabeth et al., 2016). Individuals who lose the hearing before acquiring speech and language, such as those with congenital deafness, are at a much greater disadvantage than those with acquired deafness in the interdependent language processes such as: thought development, concepts in number, measurement, operations, problem solving, and so on (Pagliaro and Kritzer, 2013; Pénicaud et al., 2013).

According to the Triple Code Model (TCM; Dehaene, 1992), a model of numerical processing proposes that numbers are represented in three codes: analog magnitude representation, auditory verbal representation, and visual Arabic representation. Dehaene and Cohen (1995, 1997) proposed two major transcoding paths between the three representational codes: a direct a semantic route that transcodes written numerals to auditory verbal to guide retrieval of rote knowledge of arithmetic facts without semantic mediation, and an indirect semantic route specialized for quantitative processing that manipulates analog magnitude representations by manipulating visual Arabic representations. Neuropsychological studies found some patients demonstrated impairment in tasks involving verbal representations of number, but could perform tasks involving non-verbal representations of number (Cipolotti and Butterworth, 1995; Cohen et al., 2000).

Although the time of onset of hearing loss is known to be an important factor influencing the academic performance of deaf individuals (Moores, 1985; Paul and Quigley, 1990; Liu, 2013), little research has focused on the arithmetical abilities of individuals with congenital or acquired deafness. Moreover, many studies focused on gender differences and mathematical performance in hearing population found male advantage in mathematics (Burton and Lewis, 1996; Gallagher et al., 2000; Perie et al., 2005; Liu and Wilson, 2009), while some showed that girls outperformed boys in numerical magnitude processing (Wei et al., 2012) and arithmetic computation (Linn and Hyde, 1989; Willingham and Cole, 1997; Wei et al., 2012), others revealed no gender differences in children's mathematical ability (Kersey et al., 2018; Zhang et al., 2020).

Therefore, the first aim of the present study is to investigate the presence of differences among the performance of hearing adolescents, adolescents with congenital and acquired deafness in the tasks of numerical magnitude comparison and arithmetic computation; a second aim is to explore the gender differences in numerical magnitude processing and arithmetic computation of deaf adolescents. And the third aim is to examine the predictive role of numerical magnitude processing on arithmetical abilities of deaf adolescents.

Numerical Magnitude Processing and Mathematical Ability

Numerical magnitude processing, as the mental manipulation of quantitative information of either symbolic numbers (e.g., Arabic digits) or non-symbolic quantities (e.g., dot arrays) (Turconi et al., 2004; Tudusciuc and Nieder, 2007) has been found important for successful mathematical development (e.g., Butterworth et al., 2011, for a review) and positively associated with mathematical performance of the hearing individuals (Sasanguie et al., 2013; Fazio et al., 2014; Schneider et al., 2017). Further research by Zhang et al. (2016) found that numerical magnitude processing was the independent predictor of arithmetical computation but not mathematical reasoning for hearing children. Butterworth (2005) claimed that numerical magnitude processing was one of the reasons for dyscalculic difficulties in arithmetic.

For deaf individuals, it has also been found that ANS (approximate number system) acuity (non-symbolic magnitude processing) is significantly associated with mathematical performance, and less acuity in the ANS, compared to hearing peers, may be the reason for their delays in mathematics achievement (Bull et al., 2006, 2018). Some studies showed significant differences between deaf and hearing individuals in response times for numerical magnitude comparison (Epstein et al., 1995; Marschark et al., 2003). However, other researchers found no significant differences between deaf and hearing individuals in their number representation processes (Zarfaty et al., 2004; Arfé et al., 2011; Barbosa, 2013). Whether numerical magnitude processing (symbolic, non-symbolic) is a predictor of the arithmetical ability of the deaf population or not still needs to be verified.

Numerical Magnitude Processing in Hearing-Impaired and Deaf Individuals

Numerical magnitude processing, or numerical magnitude representation process, in hearing-impaired and deaf individuals has been analyzed in both children and adults. Zarfaty et al. (2004) compared 3- and 4-year-old deaf and hearing children's performance in number representation tasks and found out the better performance of deaf children in the spatial task and no difference from hearing counterparts in the temporal tasks. Barbosa (2013) conducted a similar study with Brazilian deaf children aged 5–6 years, and found out the young deaf children's number representation ability was as good as that of hearing children, which supported the previous research findings by Zarfaty et al. (2004). Arfé et al. (2011) investigated number representation ability of deaf primary school children with cochlear implants in a digit comparison task and an analogic comparison task, and also found out the better performance of deaf children in the analogic task and no difference from hearing children in the digit comparison task. All of these studies, on symbolic magnitude processing, confirmed that deaf children present the same abilities in number representation as their hearing peers.

Bull et al. (2005) investigated deaf adults' performance on a magnitude comparison task, and found that deaf participants performed more slowly than hearing participants in making comparative judgments. However, there was no substantial difference in the basic numerical magnitude processing capacity between the deaf and age-matched hearing peers. Rodríguez-Santos et al. (2014) explored deaf and hard-of-hearing children's numerical magnitude representation process by means of symbolic (Arabic digits) and non-symbolic (dot constellations) magnitude comparison tasks, and found out slower reaction times of deaf participants in the symbolic but not non-symbolic task, which was believed to the delay that deaf individuals experienced in accessing representations from symbolic codes. Bull et al. (2018) also found that children with hearing loss had poorer numerical discrimination skills and less acuity in the ANS (non-symbolic magnitude processing) compared to hearing peers. As we can see, previous studies have rarely examined both the symbolic and non-symbolic magnitude processing of deaf individuals at the same time, except for Rodríguez-Santos et al. (2014), and diverged in whether they have the similar numerical magnitude representation to their hearing peers.

The Present Study

As language abilities can support mathematical performance in deaf individuals (Kelly and Gaustad, 2007; Andin et al., 2014; Huber et al., 2014; Vitova et al., 2014) and some studies have showed a link between sign language skills and reading ability in deaf individuals (Mayberry et al., 2011; Rudner et al., 2012), reading comprehension has been used as a task to evaluate the linguistic performance for deaf adolescents, and as a control variable in this study. And according to previous studies, the general cognitive abilities (i.e., non-verbal IQ, processing speed) of deaf students affect their mathematical performance (Chen et al., 2019; Chen and Wang, 2020), so we also take the non-verbal IQ and processing speed as the control variables.

The aim of the present study was to examine numerical magnitude processing and its contribution to arithmetical ability in deaf adolescents. Firstly, according to the TCM, indirect semantic route supports the non-verbal numerical magnitude processing that manipulates analog magnitude representations by manipulating visual Arabic representations. Compared to the acquired deafness, individuals with congenital deafness may be more dependent on this non-verbal numerical magnitude processing due to auditory deprivation. Therefore, we want to explore whether there are differences across groups of congenital and acquired deafness in numerical magnitude processing and arithmetical ability.

Secondly, the researches on whether there are gender differences in mathematical performance of hearing individuals are still controversial, so we want to explore the gender differences in numerical magnitude processing and arithmetic computation of deaf adolescents. Thirdly, in view of the importance of mathematical ability and the lag of deaf children in arithmetic (Traxler, 2000; Swanwick et al., 2005; Gottardis et al., 2011), against the background of the found associations between numerical magnitude processing and arithmetical ability of hearing individuals (Fazio et al., 2014; Zhang et al., 2016; Schneider et al., 2017) and no significant differences between deaf and hearing individuals in their number representation processes (Zarfaty et al., 2004; Arfé et al., 2011; Barbosa, 2013), we also aimed to examine whether numerical magnitude processing (symbolic, non-symbolic) is a predictor of the arithmetical ability of the deaf adolescents.

METHODS

Participants

The study included 58 congenital deaf adolescents [M_{age} = 184.36 (107–227) ± 28.12 months; 29 girls; $M_{unaided\ PTA\ loss\ in\ better\ ear}$ = 98.54 ± 16.45 dB, 60–120 dB; Note PTA means Pure Tone Average; In amplification: 21 use of hearing aids, 10 use of cochlear implants, 31 no use of hearing aids and cochlear implants; Mode of family communication: 35 in Mandarin sign/spoken language mix, 16 in spoken Mandarin, seven in Mandarin sign language], 54 acquired deaf adolescents[M_{age} = 188.44 (99–231) ± 26.48 months; 27 girls; $M_{unaided\ PTA\ loss\ in\ better\ ear}$ = 99.29 ± 12.72 dB, 75–110 dB; In amplification: 26 use of hearing aids, eight use of cochlear implants, 22 no use of hearing aids and cochlear implants; Mode of family communication: 47 in Mandarin sign/spoken language mix, six in spoken Mandarin, one in Mandarin sign language], and 58 hearing adolescents [M_{age} = 166.34 (98–187) ± 21.75] months; 27 girls. Deaf participants were recruited from the special education schools in the Haikou municipality of Hainan Province in China with moderate to severe hearing impairment (60–120 dB). All participants had normal or corrected-to-normal vision. Adolescents with congenital deafness, who were born with deafness, were assigned to the congenital group, adolescents with acquired deafness, whose hearing impairment was not present at birth but developed sometimes during life, were assigned to the acquired group. The congenital and acquired groups matched in age, gender, hearing loss, and intelligence; all the groups (including hearing group) matched in intelligence. The university's institutional review board approved the study. Participants' and their parents' consents were obtained prior to classroom-based testing.

Measures

Non-verbal IQ

The non-verbal matrices task, which was adapted from Raven's Progressive Matrices (Raven, 2000), was used to assess non-verbal IQ. It is a simplified version of Raven's Progressive Matrices that only had two candidate answers for each question, instead of 4–6 choices in the original version. Due to time constraints, the task was shortened to 80 items, 44 of which came from Standard Progressive Matrices (12 from the first set and eight from each of the other four sets) and 36 from Advanced Progressive Matrices. In the test, a large figure with a missing segment appeared in the center of the computer screen, and there were two options below. Participants were asked to identify the missing segment according to the rules underlying the figure, and pressed the "Q" key when the missing segment was on the left or the "P" key when it was on the right.

Processing Speed

A simple reaction time task was used to measure the processing speed [cf., Butterworth's (2003) "Dyscalculia Screener," which included a reaction time task]. Each trial presented a fixation "+" in the center of the black computer screen, and a white dot appeared at the 30 degree angle randomly on the left or right side of the fixation "+." Participants were asked to press the "Q" key when the white dot appeared on the left or the "P" key when the white dot appeared on the right. There were 30 trials in the test, of which 15 were white dots on the left and 15 were white dots on the right side of the fixation "+." The dots were randomly presented, and the interval between responses and stimuli was varied randomly between 1000 and 2000 ms.

Reading Comprehension

The sentence completion task, which was adapted from Siegel and Ryan (1988), was used to measure reading comprehension (Elbeheri et al., 2011; Träff et al., 2018; Cui et al., 2019). Materials for the task were selected from the test materials used in primary and middle schools in China (from first to ninth grade). On the test, a sentence was presented in the center of the computer screen with a word missing and there were two options below. Participants were asked to choose a word from the options to complete the sentence and press the "Q" key if the correct answer was on the left, or press the "P" key if the correct answer was on the right. There were 120 problems on the test, ordered from easy to difficult, and the time interval for each problem was 1000 ms.

Numerical Magnitude Comparison

Symbolic Magnitude Comparison

A classic numerical magnitude comparison task, which was adapted from Zhou et al. (2007), used a Stroop-like paradigm to measure the ability to compare numerical values of numbers that varied in physical size (1:2 size ratio). In this task, participants had to indicate the numerically larger of two simultaneously presented Arabic digits (ranging from 2 to 9), one displayed on the left and the other on the right side of the computer screen in random orders, ignoring the differences in physical size. The position of the largest number was counterbalanced. There were 84 trials, and the stimulus interval was 1000 ms.

Non-symbolic Magnitude Comparison

The non-symbolic magnitude comparison task, which was used to assess approximate number sense (ANS) (e.g., Wei et al., 2012; Zhou et al., 2015), was divided into three sessions, with 40 trials in each session, and participants were required to complete all 120 trials. In this task, participants had to indicate the larger of two simultaneously presented dot arrays with different sizes and numbers, one displayed on the left and the other on the right side of the computer screen, ignoring all visual properties, such as total surface area, envelope area, diameter, and circumference. The dot arrays were created following a common procedure to control for continuous quantities in non-symbolic numerical discrimination (e.g., Halberda et al., 2008; Agrillo et al., 2013). The number of dots in each dot array varied from 5 to 32. The position of the largest numerosity was counterbalanced. The presentation time of each trial was 200 ms, and the interval time was 840 ms.

Arithmetic Computation

Simple Subtraction

The simple subtraction task, which consisted of 92 problems, was the reversed operation to single-digit addition. For each trial, a subtraction problem (e.g., 17–9) of <20 was presented at the top of the computer screen, and two candidate answers were presented on the bottom. The largest minuend of the problem was 18, and the smallest one was 2. The differences between two operands were always single-digit numbers, so the answer ranged from 2 to 9. The false candidate answer deviated from the true answer by plus or minus 1 to 3 (i.e., ± 1, ± 2, or ± 3). Participants were asked to press the "Q" key if the true answer was on the left or press the "P" key if it was on the right. This was a time-limited (2 min) task, and the interval time of each trial (problem) was 1000 ms.

Complex Subtraction

The complex subtraction task, which consisted of 95 problems, included double-digit numbers for both operands. For each trial, a subtraction problem (e.g., 82–37) of <100 was presented at the top of the computer screen, and two candidate answers were presented on the bottom. Borrowing was required for most problems. The differences between the false answers and the true answers were 1 or 10. The task was limited to 2 min, and the interval time of each problem was 1000 ms.

Procedure

All participants were tested at their own school during regular school hours and all tasks were computerized using the E-prime 2.0 software and were all administered using a 15 inch laptop individually in a quiet room. The experimenters, the teachers of the participants in the Department of Deaf, who were proficient in sign language and familiar with the specific situation of the participants, explained the instructions with slides and sign language and participants were instructed to perform both accurately and quickly by pressing the "Q" or "P" keys on a computer keyboard. Before the formal testing started, there was a practice session and feedback: When the item was correctly answered, the computer screen read "Correct! Can you go faster?" When participants answered incorrectly, the screen read "It is wrong. Try again." Each trial started with a 200 ms fixation cross in the center of the computer screen. After 1000 ms the stimuli appeared and remained visible until response, except for the non-symbolic magnitude comparison task where the stimuli disappeared after 840 ms, in order to avoid counting. Accuracy (ACC) and RT (in milliseconds) were recorded for processing speed and numerical magnitude comparison tasks. Answers and reaction times were recorded by the laptop.

In order to control for the effect of guessing, the adjusted score was used in the tests such as non-verbal IQ, reading comprehension and arithmetic computation (simple and complex subtraction). It was calculated by subtracting the number of incorrect responses from the number of correct responses following the Guilford correction formula "S = R –

W/(n – 1)" (S: the adjusted number of items that the participants can actually perform without the aid of chance. R: the number of right responses, W: the number of wrong responses. n: the number of alternative responses to each item) (Guilford, 1936). This correction procedure has been utilized recently in studies of mathematical cognition (Cirino, 2011; Zhou et al., 2015; Cui et al., 2019).

Statistical Analyses

The statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS, version 25.0). Descriptive statistics were computed for demographic data and all study variables. One-way analyses of variance (ANOVAs) and LSD *post-hoc* comparisons were carried out to compare the differences in all the measures on the study groups. The repeated measurement analyses of variance (ANOVAs), with the group (congenital deaf adolescents, acquired deaf adolescents, hearing adolescents) and gender as between-subject factors and mathematical tasks as within-subject factors, were conducted to analyze group differences for accuracy and reaction times in the two numerical magnitude comparison tasks and the scores in arithmetic computation tasks. In order to control the effect of general cognitive abilities (e.g., reading comprehension) on mathematical tests, we used non-verbal IQ, processing speed and reading comprehension as covariates for ANOVAs. Pearson's correlation coefficients were calculated between the scores of all cognitive and mathematical tests. A series of linear hierarchical regression analyses were conducted to test the role of numerical magnitude processing (symbolic and non-symbolic numerical magnitude comparison) to arithmetic computation (simple and complex subtraction) of deaf adolescents, while controlling for demographic variables (i.e., age, gender; entering stage 1) and general cognitive abilities (i.e., non-verbal IQ, processing speed and reading comprehension; entering stage 2).

RESULTS

Descriptive Statistics

The means and standard deviations and one-way analyses of variance of the scores for all seven tasks on the study groups are displayed in **Table 1**. We found a significant group effect on reading comprehension, arithmetic computation (simple and complex subtraction), symbolic magnitude comparison (accuracy and reaction time), and the accuracy of non-symbolic magnitude comparison but not on the reaction time of non-symbolic magnitude comparison. Hearing group outperformed congenital group, and congenital group outperformed acquired group in arithmetic computation and symbolic magnitude comparison (reaction time).

Numerical Magnitude Comparison

A 2 × 2 ×3 mixed model, repeated measures ANOVA was conducted to examine whether the accuracy of numerical magnitude processing (symbolic, non-symbolic) varied by gender and group (see **Figure 1**). There was one within-subjects factor (numerical magnitude comparison: symbolic vs. non-symbolic) and two between-subjects factors: (gender: boys vs. girls) and (group: congenital, acquired, hearing). In order to control the effect of general cognitive abilities (e.g., reading comprehension) on numerical magnitude comparison, we used non-verbal IQ, processing speed and reading comprehension as covariates for ANOVA.

The main effect of gender, $F_{(1,160)} = 5.26$, $\eta^2 = 0.03$, $p < 0.05$, was significant, indicating that the numerical magnitude processing of boys (81.00 ± 1.00%) was more accurate than that of girls (77.71 ± 1.02%). There was no a main effect of numerical magnitude comparison, $F_{(1,160)} = 0.20$, $\eta^2 = 0.001$, $p > 0.05$, and there was no a main effect of group, $F_{(2,160)} = 0.92$, $\eta^2 = 0.011$, $p > 0.05$; but the group × gender interaction was significant, $F_{(2,160)} = 5.05$, $\eta^2 = 0.06$, $p < 0.01$. The simple effect test showed that for boys, there were no significant differences among the three groups ($p > 0.05$); for girls, there was no significant difference between congenital group and hearing group ($p > 0.05$), but the scores of acquired group were lower than those of hearing group significantly ($p < 0.01$) and congenital group marginally significantly ($p = 0.056$).

There were no significant two-way numerical magnitude comparison × group interaction, $F_{(2,160)} = 0.56$, $\eta^2 = 0.007$, $p = 0.57$, and numerical magnitude comparison × gender interaction, $F_{(1,160)} = 0.01$, $\eta^2 = 0.000$, $p = 0.92$. And there was no significant three-way numerical magnitude comparison × group × gender interaction, $F_{(2,160)} = 0.17$, $\eta^2 = 0.002$, $p = 0.84$.

In order to examine whether the reaction time of numerical magnitude processing (symbolic, non-symbolic) varied by gender and group, a 2 × 2 × 3 mixed model, repeated measures ANOVA was again conducted with numerical magnitude comparison (symbolic vs. non-symbolic) as within-subject factor, gender (boys vs. girls), and group (congenital, acquired, hearing) as between-subject factors, and general cognitive abilities (non-verbal IQ, processing speed and reading comprehension) as covariates (see **Figure 2**).

The main effect of group, $F_{(2,160)} = 3.42$, $\eta^2 = 0.04$, $p < 0.05$, was significant; LSD *post-hoc* comparisons showed that there was no significant difference in the reaction times between the congenital group and the hearing group ($p > 0.05$), but the reaction times of the acquired group were significantly longer than those of the congenital group and the hearing group ($p < 0.05$). There was no a main effect of numerical magnitude comparison, $F_{(1,160)} = 0.57$, $\eta^2 = 0.004$, $p > 0.05$, and there was no a main effect of gender, $F_{(1,160)} = 0.02$, $\eta^2 = 0.000$, $p > 0.05$; but the numerical magnitude comparison × gender interaction was significant, $F_{(1,160)} = 7.29$, $\eta^2 = 0.04$, $p < 0.01$. The simple effect test showed that for boys and girls, the reaction time in symbolic magnitude comparison task was significantly longer than that in non-symbolic magnitude comparison task ($p < 0.001$).

There were no significant two-way numerical magnitude comparison × group interaction, $F_{(2,160)} = 0.51$, $\eta^2 = 0.006$, $p = 0.60$, and group × gender interaction, $F_{(2,160)} = 0.01$, $\eta^2 = 0.000$, $p = 0.99$. And there was no significant three-way numerical magnitude comparison × group × gender interaction, $F_{(2,160)} = 1.32$, $\eta^2 = 0.016$, $p = 0.27$.

TABLE 1 | All the measures on the study groups (M ± SD).

	Index	A. Congenital group	B. Acquired group	C. Hearing group	Minimum	Maximum	Statistical difference	
		(*n* = 58)	(*n* = 54)	(*n* = 58)			*F(2, 167)*	*LSD*
Age (months)		184.36 ± 28.12	188.44 ± 26.48	166.34 ± 21.75	98	231	12.03***	B, A > C
Non-verbal IQ	Adj. No. of correct response	12.36 ± 10.01	12.39 ± 10.71	12.79 ± 10.31	−12	30	0.03	—
PS. (ACC)	Accuracy (%)	93.55 ± 13.22	93.39 ± 13.38	95.73 ± 8.92	46	100	0.68	—
PS. (RT)	Reaction time (Millisecond)	483.56 ± 137.56	511.79 ± 135.35	411.49 ± 159.48	230.75	1232.50	7.23**	B, A > C
Reading Com.	Adj. No. of correct response	8.00 ± 10.83	4.46 ± 9.02	30.78 ± 8.48	−23	47	128.60***	B, A < C
Symbolic (ACC)	Accuracy (%)	87.41 ± 13.79	85.07 ± 16.82	93.72 ± 6.13	48	100	6.77**	B, A < C
Symbolic (RT)	Reaction time (Millisecond)	658.61 ± 145.92	719.95 ± 170.82	582.73 ± 120.24	393.00	1159.00	12.35***	B > A > C
Non-symbolic (ACC)	Accuracy (%)	68.00 ± 13.77	67.20 ± 15.30	74.38 ± 13.10	42	93	4.47*	B, A < C
Non-symbolic (RT)	Reaction time (Millisecond)	484.19 ± 133.38	533.10 ± 193.14	515.94 ± 136.16	232.00	1058.00	1.43	—
Simple subtraction	Adj. No. of correct response	27.97 ± 15.48	21.07 ± 14.52	41.94 ± 8.31	−5	59	37.06***	B < A < C
Complex subtraction	Adj. No. of correct response	10.67 ± 9.34	6.50 ± 9.78	19.31 ± 9.00	−19	33	27.46***	B < A < C

Adj., adjusted; No., number; PS., Processing speed; Reading Com., Reading comprehension; ACC, accuracy; RT, reaction time.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

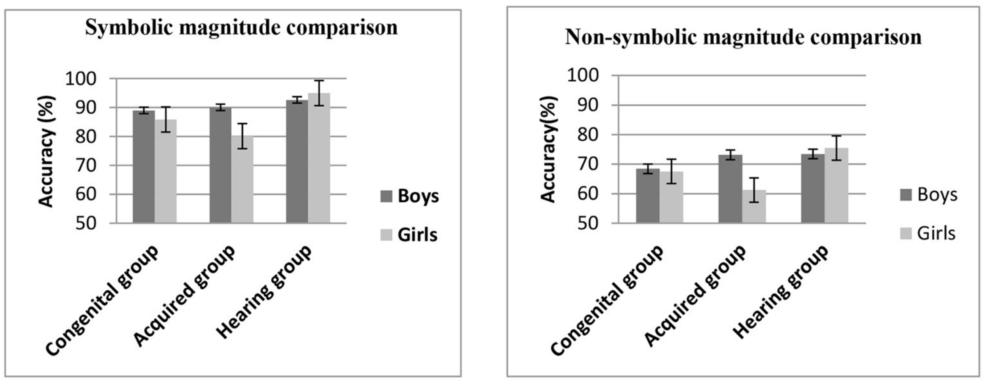


FIGURE 1 | Mean accuracy (%) on the numerical magnitude comparison task (symbolic and non-symbolic) across groups.

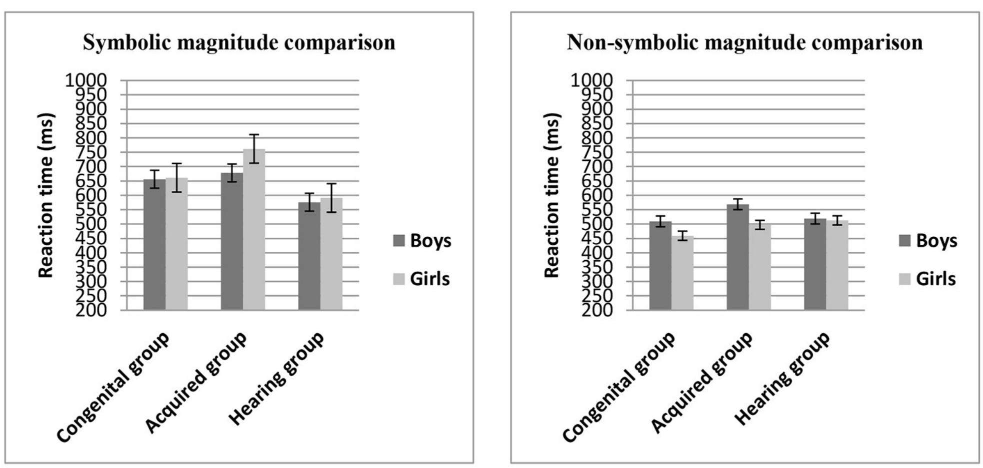


FIGURE 2 | Mean reaction time (ms) on the numerical magnitude comparison task (symbolic and non-symbolic) across groups.

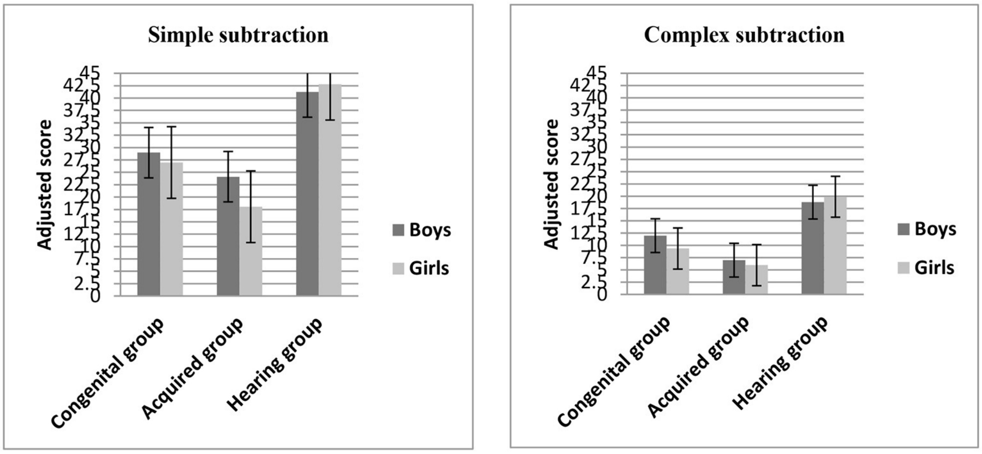


FIGURE 3 | Mean adjusted scores on the arithmetic computation tasks (simple and complex subtraction) across groups.

Arithmetic Computation

To examine whether the performance of arithmetic computation (simple and complex subtraction) varied by gender and group, a 2 × 2 × 3 mixed model, repeated measures ANOVA was again conducted with arithmetic type (simple vs. complex) as within-subject factor, gender (boys vs. girls) and group (congenital, acquired, hearing) as between-subject factors, and general cognitive abilities (non-verbal IQ, processing speed, and reading comprehension) as covariates (see **Figure 3**).

The main effect of group, $F_{(2,160)} = 3.32$, $\eta^2 = 0.04$, $p < 0.05$, was significant; LSD *post-hoc* comparisons showed that there were no significant differences in the scores between the congenital group and the hearing group ($p > 0.05$), but the scores of the acquired group were significantly lower than those of the congenital group and the hearing group ($p < 0.05$). There was no a main effect of arithmetic type, $F_{(1,160)} = 0.06$, $\eta^2 = 0.000$, $p > 0.05$, and there was no a main effect of gender, $F_{(1,160)} = 2.11$, $\eta^2 = 0.013$, $p > 0.05$.

And there were no significant two-way arithmetic type × group interaction, $F_{(2,160)} = 0.47$, $\eta^2 = 0.006$, $p > 0.05$; arithmetic type × gender interaction, $F_{(1,160)} = 1.61$, $\eta^2 = 0.01$, $p > 0.05$, and group × gender interaction, $F_{(2,160)} = 0.40$, $\eta^2 = 0.005$, $p > 0.05$. And there was no significant three-way arithmetic type × group × gender interaction, $F_{(2,160)} = 0.99$, $\eta^2 = 0.012$, $p > 0.05$.

Numerical Magnitude Comparison and Arithmetic Computation

In order to explore the numerical magnitude processing in deaf adolescents and its contribution to arithmetical ability, we first analyze how numerical magnitude comparison and arithmetic computation may differ across groups and then consider the contribution of numerical magnitude processing to arithmetical ability within each group and in all deaf adolescents.

Analysis of Each Group of Deaf Adolescents

The partial correlations were separately calculated for each group between arithmetic computation (simple and complex subtraction), general cognitive abilities (i.e., non-verbal IQ, processing speed, and reading comprehension), and numerical magnitude processing (symbolic and non-symbolic) in deaf adolescents. A Bonferroni correction was used to maintain the p-value < 0.05 across the 45 correlations in **Tables 2**, **3**. Thus, a conservative p-value of <0.00111 ($=0.05/45$) was considered statistically significant. As shown in **Tables 2**, **3**, there was only a significant correlation between reading comprehension and simple subtraction in congenital group; However, there was a significant correlation between the accuracy of numerical magnitude processing (symbolic and non-symbolic) and simple subtraction and a significant correlation between the accuracy of non-symbolic numerical magnitude processing and complex subtraction, except for the significant correlation between reading comprehension and simple subtraction, in acquired group.

A series of linear hierarchical regression analyses were conducted separately for each group to determine the contribution of numerical magnitude processing to the arithmetic ability (simple and complex subtraction) of deaf adolescents within each group. We also performed Bonferroni correction on the 2 regression analyses. Thus, a conservative p-value of <0.025 ($=0.05/2$) was considered statistically significant. According to **Table 4**, except that general cognitive abilities could account for 27.3% of the variation in simple subtraction [$F_{\text{change }(4,51)} = 5.28$, $p = 0.001$] and demographic variables could account for 14.1% of the variation in complex subtraction [$F_{\text{change }(2,55)} = 4.51$, $p = 0.015$], others did not have a contribution to the arithmetic ability of deaf adolescents in congenital group. However, general cognitive abilities [$F_{\text{change }(4,47)} = 4.95$, $p = 0.002$] could account for 27.9% and symbolic magnitude processing [$F_{\text{change }(2,45)} = 8.98$, $p = 0.001$] could account for 18.9% of the variation in simple subtraction; and general cognitive abilities [$F_{\text{change }(4,47)} =$

TABLE 2 | Partial correlations after controlling for age and gender differences among all the test scores in congenital group.

	1	2	3	4	5	6	7	8	9	10
1 Non-verbal IQ	–									
2 PS. (ACC)	0.11	–								
3 PS. (RT)	−0.11	−0.29	–							
4 Reading Com.	0.35	0.07	−0.37	–						
5 Symbolic (ACC)	0.34	0.42*	−0.25	0.14	–					
6 Symbolic (RT)	−0.11	−0.13	0.33	−0.16	−0.08	–				
7 Non-symbolic (ACC)	0.15	0.11	−0.10	0.31	0.28	−0.37	–			
8 Non-symbolic (RT)	0.09	−0.09	−0.08	0.16	0.30	−0.08	0.73*	–		
9 Simple subtraction	0.29	0.26	−0.34	0.46*	0.30	−0.32	0.32	0.11	–	
10 Complex subtraction	0.21	0.24	−0.22	0.27	0.24	−0.31	0.30	0.12	0.66*	–

**p* < 0.05, Bonferroni-corrected. PS., Processing speed; Reading Com., Reading comprehension; ACC, accuracy; RT, reaction time.*

TABLE 3 | Partial correlations after controlling for age and gender differences among all the test scores in acquired group.

	1	2	3	4	5	6	7	8	9	10
1 Non-verbal IQ	–									
2 PS. (ACC)	0.37	–								
3 PS. (RT)	−0.22	−0.17	–							
4 Reading Com.	0.39	0.20	−0.35	–						
5 Symbolic (ACC)	0.35	0.26	−0.18	0.24	–					
6 Symbolic (RT)	−0.06	−0.15	0.44*	−0.20	−0.02	–				
7 Non-symbolic (ACC)	0.31	0.18	−0.39	0.31	0.31	−0.29	–			
8 Non-symbolic (RT)	0.21	0.21	0.00	0.15	0.10	0.33	0.53*	–		
9 Simple subtraction	0.30	0.17	−0.34	0.51*	0.51*	−0.37	0.48*	0.11	–	
10 Complex subtraction	0.25	−0.10	−0.27	0.34	0.25	−0.25	0.55*	0.31	0.56*	–

**p* < 0.05, Bonferroni-corrected. PS., Processing speed; Reading Com., Reading comprehension; ACC, accuracy; RT, reaction time.*

3.05, $p = 0.026$] could account for 19.6% and non-symbolic magnitude processing [$F_{\text{change}\ (2,43)} = 7.12$, $p = 0.002$] could account for 17.3% of the variation in complex subtraction in acquired group.

Analysis of All Deaf Adolescents

To examine the association between numerical magnitude processing and arithmetic ability in all deaf adolescents, partial correlations were computed and a Bonferroni correction was also used to maintain the p-value < 0.05 across the 45 correlations in **Table 5**. Thus, a conservative p-value of <0.00111 (=0.05/45) was considered statistically significant. As shown in **Table 5**, there was a significant correlation between deaf adolescents' reaction time on the symbolic magnitude comparison task and their performance on the arithmetic computation tasks (simple and complex subtraction), and there was also a significant correlation between deaf adolescents' accuracy on the non-symbolic magnitude comparison task and their performance on the arithmetic computation tasks (simple and complex subtraction).

In order to determine the contribution of numerical magnitude processing to the arithmetic ability of all deaf adolescents, a series of linear hierarchical regression analyses were conducted. We also performed Bonferroni correction on the two regression analyses. Thus, a conservative p-value of <0.025 (=0.05/2) was considered statistically significant. According to **Table 6**, general cognitive abilities could account for 27.3% of the variation in simple subtraction [$F_{\text{change}\ (4,104)} = 10.86$, $p < 0.001$]. After controlling for scores of general cognitive ability and demographic variables, symbolic magnitude processing could account for 9.8% of the variation in simple subtraction [$F_{\text{change}\ (2,102)} = 8.95$, $p < 0.001$]. However, demographic variables [$R^2 = 0.072$, $F_{\text{change}\ (3,108)} = 2.81$, $p = 0.043$, $p > 0.025$] and non-symbolic magnitude processing [$R^2 = 0.021$, $F_{\text{change}\ (2,100)} = 1.94$, $p = 0.149$] did not have an additional contribution to simple subtraction.

Demographic variables (age, gender, onset of hearing loss) could account for 12.1% of the variation [$F_{\text{change}\ (3,108)} = 4.97$, $p < 0.01$], and general cognitive abilities could account for 11.0% of the variation [$F_{\text{change}\ (4,104)} = 3.71$, $p < 0.01$] in complex subtraction. After controlling for scores of general cognitive abilities and demographic variables, symbolic magnitude processing could account for 5.3% of the variation [$F_{\text{change}\ (2,102)} = 3.81$, $p < 0.025$] and non-symbolic magnitude processing could account for 5.4% of the variation [$F_{\text{change}\ (2,100)} = 4.11$, $p < 0.025$] in complex subtraction.

TABLE 4 | Hierarchical regression models predicting arithmetic ability (simple and complex subtraction) from age, gender, general cognitive ability, symbolic, and non-symbolic magnitude processing in congenital and acquired group.

	Simple subtraction β				Complex subtraction β			
	Step 1	**Step2**	**Step3**	**Step4**	**Step 1**	**Step2**	**Step3**	**Step4**
Congenital group								
Age (months)	0.252	0.184	0.132	0.066	0.348*	0.297	0.234	0.173
Gender	−0.064	−0.096	−0.068	−0.093	−0.138	−0.154	−0.128	−0.141
Non-verbal IQ	–	0.136	0.079	0.081	–	0.117	0.070	0.077
Ps. (ACC)	–	0.176	0.120	0.084	–	0.173	0.130	0.115
Ps. (RT)	–	−0.147	−0.066	−0.105	–	−0.089	−0.003	−0.037
Reading Com.	–	0.333*	0.334*	0.292	–	0.164	0.164	0.118
Symbolic (ACC)	–	–	0.142	0.147	–	–	0.101	0.080
Symbolic (RT)	–	–	−0.210	−0.139	–	–	−0.247	−0.181
Non-symbolic (ACC)	–	–	–	0.229	–	–	–	0.197
Non-symbolic (RT)	–	–	–	−0.157	–	–	–	−0.073
	$R^2 = 0.068$	$R^2 = 0.341$*	$R^2 = 0.388$	$R^2 = 0.401$	$R^2 = 0.141$*	$R^2 = 0.259$	$R^2 = 0.312$	$R^2 = 0.325$
		($\Delta R^2 = 0.273$*)	($\Delta R^2 = 0.047$)	($\Delta R^2 = 0.013$)		($\Delta R^2 = 0.118$)	($\Delta R^2 = 0.053$)	($\Delta R^2 = 0.012$)
Acquired group								
Age (months)	−0.117	−0.103	−0.216	−0.198	0.219	0.176	0.103	0.141
Gender	−0.198	−0.167	0.016	0.071	−0.077	−0.025	0.078	0.218
Non-verbal IQ	–	0.101	0.027	−0.005	–	0.207	0.178	0.111
Ps. (ACC)	–	0.028	−0.059	−0.062	–	−0.256	−0.303	−0.356*
Ps. (RT)	–	−0.170	−0.020	0.026	–	−0.177	−0.077	0.005
Reading Com.	–	0.389*	0.330*	0.310*	–	0.238	0.205	0.147
Symbolic (ACC)	–	–	0.428*	0.390*	–	–	0.201	0.157
Symbolic (RT)	–	–	−0.307*	−0.283	–	–	−0.218	−0.298
Non-symbolic (ACC)	–	–	–	0.209	–	–	–	0.277
Non-symbolic (RT)	–	–	–	0.022	–	–	–	0.279
	$R^2 = 0.058$	$R^2 = 0.337$*	$R^2 = 0.526$*	$R^2 = 0.558$	$R^2 = 0.050$	$R^2 = 0.246$*	$R^2 = 0.304$	$R^2 = 0.477$*
		($\Delta R^2 = 0.279$*)	($\Delta R^2 = 0.189$*)	($\Delta R^2 = 0.032$)		($\Delta R^2 = 0.196$*)	($\Delta R^2 = 0.058$)	($\Delta R^2 = 0.173$*)

$p < 0.05$, Bonferroni-corrected. Ps., Processing speed; Reading Com., Reading comprehension; ACC, accuracy; RT, reaction time.

DISCUSSION

The current study aimed to examine numerical magnitude processing and its contribution to arithmetical ability in deaf adolescents. The main results are summarized as follows: First, repeated measures ANOVA showed that the numerical magnitude processing of boys was more accurate than that of girls. For boys, there were no significant differences among the three groups (congenital, acquired, and hearing) in the accuracy of numerical magnitude processing; for girls, there was no significant difference between congenital group and hearing group, but the accuracy in acquired group was lower than that in hearing and congenital group significantly. Second, one-way ANOVA showed hearing adolescents outperformed deaf adolescents in arithmetic computation (simple and complex subtraction), symbolic magnitude processing (accuracy and reaction time), and the accuracy, but not the reaction time of non-symbolic magnitude processing. Third, the hierarchical regression analyses of each group of deaf adolescents showed that numerical magnitude processing did not have a contribution to arithmetic computation in congenital group, but symbolic magnitude processing could contribute to simple subtraction and non-symbolic magnitude processing could contribute to complex subtraction in acquired group.

Numerical Magnitude Processing and Arithmetic Ability in Deaf Adolescents

The results of one-way ANOVA showed that deaf adolescents lag behind hearing adolescents in arithmetic computation (simple and complex subtraction), symbolic magnitude processing (accuracy and reaction time), and the accuracy, but not the reaction time of non-symbolic magnitude processing. It is basically consistent with the previous results (e.g., Rodríguez-Santos et al., 2014; Masataka, 2006) that deaf individuals were found worse performance on symbolic but not non-symbolic magnitude processing, indicating the delay of deaf individuals in symbolic but not non-symbolic encoding. According to "access deficit hypothesis" (Rouselle and Noël, 2007), deficits in the representation of numerical information in long-term memory are not general, but are linked to the numerical representation codes used for its acquisition (Arabic numerals, number words). Deaf individuals' poor performance on an Arabic number comparison task, but not on a dot collection comparison task, could be explained by difficulties in accessing the semantic

TABLE 5 | Partial correlations after controlling for age and gender differences among all the test scores.

	1	2	3	4	5	6	7	8	9	10
1 Non-verbal IQ	—									
2 PS. (ACC)	0.26	—								
3 PS. (RT)	−0.17	−0.23	—							
4 Reading Com.	0.36*	0.12	−0.37*	—						
5 Symbolic (ACC)	0.34*	0.34*	−0.24	0.20	—					
6 Symbolic (RT)	−0.08	−0.13	0.42*	−0.21	−0.08	—				
7 Non-symbolic (ACC)	0.27	0.17	−0.29	0.30*	0.31*	−0.34*	—			
8 Non-symbolic (RT)	0.18	0.10	−0.02	0.12	0.17	0.19	0.60*	—		
9 Simple subtraction	0.31*	0.23	−0.36*	0.49*	0.41*	−0.37*	0.42*	0.10	—	
10 Complex subtraction	0.23	0.07	−0.26	0.32*	0.25	−0.31*	0.42*	0.19	0.63*	—

$p < 0.05$, Bonferroni-corrected. PS., Processing speed; Reading Com., Reading comprehension; ACC, accuracy; RT, reaction time.

TABLE 6 | Hierarchical regression models predicting arithmetic ability (simple and complex subtraction) from age, gender, group, general cognitive ability, symbolic, and non-symbolic magnitude processing.

	Simple subtraction				Complex subtraction			
	Step 1	Step2	Step3	Step4	Step 1	Step2	Step3	Step4
	β	β	β	β	β	β	β	β
Age (months)	0.069	0.044	−0.025	−0.044	0.258*	0.219*	0.158	0.145
Gender	−0.133	−0.129	−0.040	−0.026	−0.105	−0.090	−0.025	0.014
Group(congenital/acquired)	−0.231*	−0.151	−0.093	−0.097	−0.234*	−0.186*	−0.138	−0.163
Non-verbal IQ	—	0.137	0.068	0.054	—	0.146	0.106	0.075
Ps. (ACC)	—	0.112	0.034	0.039	—	−0.029	−0.078	−0.079
Ps. (RT)	—	−0.174	−0.052	−0.043	—	−0.151	−0.051	−0.037
Reading Com.	—	0.335*	0.322*	0.297*	—	0.174	0.164	0.123
Symbolic (ACC)	—	—	0.280*	0.241*	—	—	0.165	0.110
Symbolic (RT)	—	—	−0.246*	−0.166	—	—	−0.223*	−0.175
Non-symbolic (ACC)	—	—	—	0.222	—	—	—	0.223
Non-symbolic (RT)	—	—	—	−0.076	—	—	—	0.068
	$R^2 = 0.072$	$R^2 = 0.346$*	$R^2 = 0.443$*	$R^2 = 0.464$	$R^2 = 0.121$*	$R^2 = 0.231$*	$R^2 = 0.284$*	$R^2 = 0.339$*
	($\Delta R^2 = 0.273$*)($\Delta R^2 = 0.098$*)($\Delta R^2 = 0.021$)				($\Delta R^2 = 0.110$*)($\Delta R^2 = 0.053$*)($\Delta R^2 = 0.054$*)			

$p < 0.05$, Bonferroni-corrected. Ps., Processing speed; Reading Com., Reading comprehension; ACC, accuracy; RT, reaction time.

information of numbers by means of symbols, due to their low-level language and their limited experience with numbers (Gregory, 1998; Nunes, 2004; Kritzer, 2009; Bull et al., 2011).

It was also found that boys outperformed girls in the accuracy of numerical magnitude processing in the study. The result was similar to the previous study of Krinzinger et al. (2012), but scare previous researches on gender differences in numerical magnitude processing in deaf individuals. Krinzinger et al. (2012) applied structural equation modeling to a longitudinal dataset of 140 primary school children and found superiority for primary school boys in numerical magnitude processing. One explanation of Krinzinger et al.'s results is that general visual-spatial abilities (but not visual-spatial working memory), which has been found to favor males (Goldstein et al., 1990; Vederhus and Krekling, 1996).

And hierarchical regression analyses of all deaf adolescents showed that numerical magnitude processing had an independent contribution to arithmetic computation after controlling for general cognitive ability. Previous studies have shown that the understanding of numerical magnitudes is helpful to the solution of arithmetic problems (De Smedt et al., 2009; Tavakoli, 2016). Neuroimaging studies have also revealed that numerical magnitude processing is related to arithmetic problem solving (Bugden et al., 2012; Price et al., 2013). It was also found the close relationship between the approximate number system acuity (non-symbolic numerical magnitude processing) and math achievement in children with hearing loss in the research of Bull et al. (2018), which is basically consistent with the results of this study.

Onset of Hearing Loss (Congenital vs. Acquired) and Mathematical Cognition of Deaf Adolescents

Acquired deafness, as the type of deafness occurring after the acquisition of speech (Hindley and Kitson, 2000), is the loss of hearing that occurs after birth and develops sometimes during a person's life. Congenital deafness, in which auditory system has not been programmed for language and communication, is the loss of hearing that was present at birth. Although the difference between congenital deafness and acquired deafness is obvious, there are few studies on the difference between them in academic achievement such as mathematics performance. DeLeon et al. (1979) explored the reading and math skills of two groups of adults either congenital or acquired deafness matched in intelligence, education level and degree of loss, and found no significant differences on reading level between the two groups, but a significantly higher math level in the congenital group than the acquired. But in the research of Ogundiran and Olaosun (2013), no significant differences were found in the academic achievement including mathematics and English Language performance between students with congenital deafness and those with acquired deafness. And the results of our study that there was no significant difference between the congenital group and the hearing group, but congenital group outperformed acquired group in numerical magnitude processing (RT) and arithmetic computation, suggesting that the mathematical cognitive abilities of the congenital deaf are better than those of the acquired deaf, which is basically consistent with the results of DeLeon et al. (1979).

Compensatory plasticity holds that the lack of auditory stimulation experienced by deaf individuals, such as congenital deafness, can be met by enhancements in visual cognition (Neville, 1990; Bavelier et al., 2006). Previous studies have shown that auditory deprivation, such as congenital deafness, can lead to enhanced peripheral visual processing, which should be contributed by the neuroplasticity in multiple systems including primary auditory cortex, supramodal, and multisensory regions (Bavelier and Neville, 2002; Scott et al., 2014). According to the connectome model, congenital sensory loss, such as congenital deafness, is thought to be a connectome disease. It is an abnormal bias in the individual wiring and coupling pattern of the brain, which might result in stronger coupling to the remaining sensory systems and reorganization within the affected sensory system. This process accounts for the abnormal visual dominance in perception after congenital deafness (Kral et al., 2016).

Although some studies have found that the processing of sign language in the brain network of congenitally deaf individuals who acquired sign language from birth from their deaf parents is similar to that for spoken words in hearing individuals. The activity in their language network is due to a kind of semantic encoding rather than visual processing (Leonard et al., 2012). The electrophysiological study of congenitally deaf adolescents revealed that better visual processing could be explained by the early latency in N1 component in visual related brain responses associated with more efficient neural processing due to auditory deprivation (Güdücü et al., 2019). And the studies of hearing individuals also showed that visual form perception had unique contributions to lower level math categories, such as numerosity comparison, digit comparison, and exact computation (Cui et al., 2017); and it was the cognitive mechanism that could explain the association between numerical magnitude processing (e.g., approximate number system) and arithmetic computation (Zhang et al., 2019).

Therefore, it may be due to the advantages of visual processing, congenitally deaf individuals outperformed acquired deafness in mathematics. And according to the TCM and related neuropsychological researches, patients (with impaired auditory speech representation) could perform non-verbal numerical magnitude processing that manipulates analog magnitude representations by manipulating visual Arabic representations (Cipolotti and Butterworth, 1995; Cohen et al., 2000). Compared to the acquired deafness, individuals with congenital deafness may be more dependent on this non-verbal, visual representation due to auditory deprivation. It is also possible because that there is only visual processing (representation) in congenital deafness, but the conversion of auditory speech and visual representation/coding is needed in acquired deafness, which may lead to the hindrance of processing.

Practical Implications

The current study offers several important insights and practical implications. First, since we found deaf adolescents lag behind hearing peers in symbolic but not non-symbolic magnitude processing, and symbolic magnitude processing accounted for unique variance in children's mathematical achievement (De Smedt et al., 2009; Bugden and Ansari, 2011), this suggests that educators should place great emphasis on helping their deaf students to understand the meaning of numerical symbols, thereby enhancing their ability to map number symbols unto non-symbolic quantities. Learning to accurately map symbolic magnitudes onto non-symbolic magnitudes is a crucial step toward performing more complex mathematics such as arithmetic operations (Siegler and Booth, 2004; Booth and Siegler, 2008; Geary et al., 2008). Second, we found general cognitive abilities (i.e., non-verbal IQ, processing speed and reading comprehension) could account for unique variance in deaf adolescents' arithmetic computation (simple and complex subtraction), which shows that the general cognitive abilities are the important influencing factors for the arithmetical ability in deaf adolescents. According to the developmental model of numerical cognition (von Aster and Shalev, 2007), the development of mathematical abilities in children is based on general cognitive abilities. Therefore, parents and teachers should promote the development of general cognitive abilities, such as intelligence, processing speed, and reading comprehension, in deaf children through activities and training as soon as possible, so as to improve their mathematics performance.

Limitations and Prospects

There are some limitations to our work. First, the sample size was limited, only 112 deaf adolescents but not young deaf children were included in this study. Second, the test of arithmetic ability only examined by simple and complex subtraction, other tests such as simple and complex addition were not included. Third, reading comprehension was only regarded as a control variable, and other language abilities were not evaluated in the present study. The neural mechanism of congenital deafness in mathematical ability should be further investigated across the groups of congenital and acquired deafness and hearing counterparts.

CONCLUSIONS

Consistent with the previous results, the study shows the worse performance on symbolic but not non-symbolic magnitude processing in deaf adolescents, which indicates that the lag of mathematics in deaf individuals may be due to the delay of symbolic but not non-symbolic encoding. It was found that boys outperformed girls in the accuracy of numerical magnitude processing in the study. Based on previous studies, it may be that the superiority of male visual-spatial ability improves their numerical magnitude processing. There was no significant difference between the congenital group and the hearing group, but congenital group outperformed acquired group in numerical magnitude processing (RT) and arithmetic computation. Similarly, it may be due to the advantage of visual processing that congenitally deaf individuals outperformed acquired deafness in mathematics. It was also found a close association between numerical magnitude processing and arithmetic computation of deaf adolescents, and after controlling for the demographic variables (age, gender, onset of hearing loss) and general cognitive ability (non-verbal IQ, processing speed, reading comprehension), numerical magnitude processing could predict arithmetic computation in all deaf adolescents but not in congenital group. The role of numerical magnitude processing (symbolic and non-symbolic) in deaf adolescents' mathematical performance should be paid attention in the training of arithmetical ability.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Hainan Normal University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

LC designed the study, carried out the experiment, analyzed the data, and wrote the manuscript. YW supervised part of the work. HW edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The work reported here was undertaken as part of a study supported by the Humanities and Social Science Fund of Ministry of Education of China [Grant Number 19YJA880002] and the Philosophy and Social Science Foundation of Hainan Province in China [Grant Number HNSK(YB)19-36]. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Agrillo, C., Piffer, L., and Adriano, A. (2013). Individual differences in non-symbolic numerical abilities predict mathematical achievements contradict ATOM. *Behav. Brain Funct*. 9:26. doi: 10.1186/1744-9081-9-26

Andin, J., Elwer, A., and Maki-Torkko, E. (2020). Arithmetic in the adult deaf signing brain. *J. Neurosci. Res.* 98, 643–654. doi: 10.1002/jnr.24569

Andin, J., Ronnberg, J., and Rudner, M. (2014). Deaf signers use phonology to do arithmetic. *Learn. Individ. Differ.* 32, 246–253. doi: 10.1016/j.lindif.2014.03.015

Arfé, B., Lucangeli, D., Genovese, E., Monzani, D., Gubernale, M., and Trevisi, P., et al. (2011). Analogic and symbolic comparison of numerosity in preschool children with cochlear implants. *Deafness Educ. Int.* 13, 34–45. doi: 10.1179/1557069X11Y.0000000002

Barbosa, H. H. (2013). Early mathematical abilities in hearing and deaf children. *Cadernos Cedes* 33, 333–347. doi: 10.1590/S0101-32622013000300003

Bavelier, D., Dye, M. W. G., and Hauser, P. C. (2006). Do deaf individuals see better? *Trends Cogn. Sci.* 10, 512–518. doi: 10.1016/j.tics.2006.09.006

Bavelier, D., and Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nat. Rev. Neurosci.* 138, 177–188. doi: 10.1038/nrn848

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

Braden, J. P. (1994). *Deafness, Deprivation, and IQ*. New York, NY: Plenum Press.

Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 35–47. doi: 10.1016/j.cognition.2010.09.005

Bugden, S., Price, G. R., McLean, D. A., and Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Dev. Cogn. Neurosci.* 2, 448–457. doi: 10.1016/j.dcn.2012.04.001

Bull, R., Blatto-Vallee, G., and Fabich, M. (2006). Subitizing, magnitude representation, and magnitude retrieval in deaf and hearing adults. *Journal of Deaf Studies and Deaf Education*, 11, 290–302. doi: 10.1093/deafed/enj038

Bull, R., Marschark, M., and Blatto-Vallee, G. (2005). SNARC hunting: examining number representation in deaf students. *Learn. Individ. Differ.* 15, 223–236. doi: 10.1016/j.lindif.2005.01.004

Bull, R., Marschark, M., Nordmann, E., Sapere, P., and Skene, W. A. (2018). The approximate number system and domain-general abilities as predictors of math ability in children with normal hearing and hearing loss. *Br. J. Dev. Psychol.* 36, 236–254. doi: 10.1111/bjdp.12204

Bull, R., Marschark, M., Sapere, R., Davidson, W. A., Murphy, D., and Nordmann, E. (2011). Numerical estimation in deaf and hearing adults. *Learn. Individ. Differ.* 27,453–457. doi: 10.1016/j.lindif.2011.02.001

Burton, N. W., and Lewis, C. (1996). Gender differences in college mathematics grades and SAT-M scores: a reanalysis of Wainer and Steinberg. *J. Educ. Meas.* 33, 257–270. doi: 10.1111/j.1745-3984.1996.tb00492.x

Butterworth, B. (2003). *Dyscalculia Screener*. London: nferNelson Pub.

Butterworth, B. (2005). The developmental of arithmetical abilities. *J. Child Psychol. Psychiatry* 46, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x

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

Chen, L., and Wang, Y. (2020). The contribution of general cognitive abilities and specific numerical abilities to mathematics achievement in students who are deaf or hard-of-hearing. *J. Dev. Phys. Disabil.* doi: 10.1007/s10882-020-09772-8. [Epub ahead of print].

Chen, L., Wang, Y., and Xiao, S. (2019). The mediating effect of intelligence between the information processing speed and numerical magnitude in hearing-impaired children. *Chin. J. Behav. Med. Brain Sci.* 38, 925–929. doi: 10.3760/cma.j.issn.1674-6554.2019.10.012

Cipolotti, L., and Butterworth, B., (1995). Toward a multiroute model of number processing: impaired number transcoding with preserved calculation skills. *J. Exp. Psychol. Gen.* 124, 375–390. doi: 10.1037/0096-3445.124.4.375

Cirino, P. T. (2011). The interrelationships of mathematical precursors in kindergarten. *J. Exp. Child Psychol.* 108, 713–733. doi: 10.1016/j.jecp.2010.11.004

Cohen, L., Dehaene, S., Chochon, F., Lehericy, S., and Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical, and fMRI study. *Neuropsychologia* 38, 1426–1440. doi: 10.1016/S0028-3932(00)00038-5

Cui, J., Zhang, Y., Cheng, D., Li, D., and Zhou, X. (2017). Visual form perception can be a cognitive correlate of lower level math categories for teenagers. *Front. Psychol.* 8:1336. doi: 10.3389/fpsyg.2017.01336

Cui, J., Zhang, Y., Wan, S., Chen, C., Zeng, J., and Zhou, X., (2019). Visual form perception is fundamental for both reading comprehension and arithmetic computation. *Cognition* 189, 141–154. doi: 10.1016/j.cognition.2019.03.014

De Smedt, B., No€ el, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? a review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001

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

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N

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

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

DeLeon, B., Berg, R., and Battin, R. (1979). Congenital vs. acquired deafness: differences in reading and mathematics levels. *J. Aud. Res.*19, 123–125.

Elbeheri, G., Everatt, J., Mahfoudhi, A., Al-Diyar, M. A., and Taibah, N. (2011). Orthographic processing and reading comprehension among Arabic speaking mainstream and LD children. *Dyslexia* 17, 123–142. doi: 10.1002/dys.430

Elizabeth, M. F., Candyce, H., Adrienne, S., Misty, P., and Eunjung, N. (2016). Sign language and spoken language for children with hearing loss: a systematic review. *Pediatrics* 137:e20151974. doi: 10.1542/peds.2015-1974

Epstein, K. I., Hillegeist, E. G., and Grafman, J. (1995). Number processing in deaf college students. *Am. Ann. Deaf* 139, 336–347. doi: 10.1353/aad.2012.0321

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

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., and Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *J. Exp. Child Psychol.* 75, 165–190. doi: 10.1006/jecp.1999.2532

Geary, D. C., Hoard, M. K., Nugent, L., and Byrd-Craven, J. (2008). Development of number-line representations in children with mathematical learning disability. *Dev. Neuropsychol.* 33, 277–299. doi: 10.1080/875656408019 82361

Goldstein, D., Haldane, D., and Mitchell, C. (1990). Sex differences in visual-spatial ability: the role of performance factors. *Mem. Cogn.* 18, 546–550. doi: 10.3758/BF03198487

Gottardis, L., Nunes, T., and Lunt, I. (2011). A synthesis of research on deaf and hearing children's mathematical achievement. *Deafness Educ. Int.* 13, 131–150. doi: 10.1179/1557069X11Y.0000000006

Gregory, S. (1998). "Mathematics and deaf children," in *Issues in Deaf Education*, eds S. Gregory, R. Knight, W. McCracken, S. Powers, and L. Watson (London: David Fulton), 119–126.

Güdücü, Ç., Ergönül, Im, Öniz, A., Ikiz, A. Ö., and Özgören, M. (2019). Deaf adolescents have bigger responses for somatosensory and visual Stimulations. *Neurosci. Lett.* 707:134283. doi: 10.1016/j.neulet.2019.134283

Guilford, J. P. (1936). The determination of item difficulty when chance success is a factor. *Psychometrika* 1, 259–264. doi: 10.1007/BF02287877

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

Hindley, D., and Kitson, N. (2000). "Mental health and deafness. London: Whurr Published Hoemann, H. W. & Briga, J. I. (1981). Hearing impairments," in *Handbook of Special Education*, eds J. M. Kauffman and D. I. Hallaham (Englewood Cliffs, NJ: Prentice-Hall), 42–74.

Huber, M., Kipman, U., and Pletzer, B. (2014). Reading instead of reasoning? predictors of arithmetic skills in children with cochlear implants. *Int. J. Pediatr. Otorhinolaryngol.* 78, 1147–1152. doi: 10.1016/j.ijporl.2014.04.038

Kelly, R. R., and Gaustad, M. G. (2007). Deaf college students' mathematical skills relative to morphological knowledge, reading level, and language proficiency. *J. Deaf Stud. Deaf Educ.* 12, 25–37. doi: 10.1093/deafed/enl012

Kennedy, C. R., Mccann, D. C., Campbell, M. J., Law, C. M., Mullee, M., and Petrou, S., et al. (2006). Language ability after early detection of permanent childhood hearing impairment. *N. Engl. J. Med.* 354, 2131–2141. doi: 10.1056/NEJMoa054915

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., and Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Sci. Learn.* 3:12. doi: 10.1038/s41539-018-0028-7

Kral, A., Kronenberger, W. G., Pisoni, D. B., and O'Donoghue, G. M. (2016). Neurocognitive factors in sensory restoration of early deafness: a connectome model. *Lancet Neurol.* 15, 610–621. doi: 10.1016/S1474-4422(16)00034-X

Krinzinger, H., Wood, G., and Willmes, K. (2012). What accounts for individual and gender differences in the multi-digit number processing of primary school children? *Z. Psychol.* 220, 78–89. doi: 10.1027/2151-2604/a000099

Kritzer, K. L. (2009). Barely started and already left behind: a descriptive analysis of the mathematics ability demonstrated by young deaf children. *J. Deaf Stud. Deaf Educ.* 14, 409–421. doi: 10.1093/deafed/enp015

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

Leonard, M. K., Ramirez, N. F., Torres, C., Travis, K. E., Hatrak, M., Mayberry, R. I., et al. (2012). Signed words in the congenitally deaf evoke typical late lexico semantic responses with no early visual responses in left superior temporal cortex. *J. Neurosci.* 32, 9700–9705. doi: 10.1523/JNEUROSCI.1002-12.2012

Linn, M., and Hyde, J. (1989). Gender, mathematics, and science. *Educ. Res.* 18, 17–27. doi: 10.3102/0013189X018008017

Linsen, S., Verschaffel, L., Reynvoet, B., and De Smedt, B., (2015). The association between numerical magnitude processing and mental versus algorithmic multi-digit subtraction in children. *Learn. Instruct.* 35, 42–50. doi: 10.1016/j.learninstruc.2014.09.003

Liu, C. (2013). *Academic and Social Adjustment Among Deaf and Hard of Hearing College Students in Taiwan*. Ann Arbor, MI: Proquest LLC.

Liu, O. L., and Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Appl. Meas. Educ.* 22, 164–184. doi: 10.1080/08957340902754635

Marschark, M., Blatto-Vallee, G., Bull, R., and Cornoldi, C. (2003). "Relative magnitude judgments by deaf and hearing individuals: surprising similarities and surprising differences," in *Paper Presented at the Annual Meeting of the Psychonomic Society* (Vancouver, BC).

Masataka, N. (2006). Differences in arithmetic subtraction of nonsymbolic numerosities by deaf and hearing adults. *J. Deaf Stud. Deaf Educ.* 11, 139–143. doi: 10.1093/deafed/enj016

Mayberry, R. I., del Giudice, A. A., and Lieberman, A. M. (2011). Reading achievement in relation to phonological coding and awareness in deaf readers: a meta-analysis. *J. Deaf Stud. Deaf Educ.* 16, 164–188. doi: 10.1093/deafed/enq049

Moores, D. F. (1985). "Educational programmes and services for hearing impaired children: issues and options," in *Education of the Hearing Impaired Child*, eds F. Powell, T. Finitzo-Hieber, S. Friel-Patti, and D. Henderson (San Diego, CA: College Hill), 3–20.

Moreno, C. (2000). *Predictors of mathematics attainment in hearing impaired children* (Unpublished PhD thesis). University of London, London, United Kingdom.

Neville, H. J. (1990). Intermodal competition and compensation in development: evidence from studies of the visual system in congenitally deaf adults. *Ann. N. Y. Acad. Sci.* 608, 71–91. doi: 10.1111/j.1749-6632.1990.tb48892.x

Nunes, T. (2004). *Teaching Mathematics to Deaf Children*. London: Whurr.

Ogundiran, O., and Olaosun, A. O. (2013). Comparison of academic achievement between students with congenital and acquired deafness in a Nigerian College. *J. Educ. Pract.* 4, 42–47.

Pagliaro, C., and Kritzer, K. (2013). The math gap: a description of the mathematics performance of preschool-aged deaf/hard-of-hearing children. *J. Deaf Stud. Deaf Educ.* 18, 139–160. doi: 10.1093/deafed/ens070

Paul, P. V., and Quigley, S. P. (1990). *Education and Deafness*. New York, NY: Longman.

Pénicaud, S., Klein, D., Zatorre, R. J., Chen, J. K., Witcher, P., Hyde, K., et al. (2013). Structural brain changes linked to delayed first language acquisition in congenitally deaf individuals. *Neuroimage* 66C, 42–49. doi: 10.1016/j.neuroimage.2012.09.076

Perie, M., Moran, R., and Lutkus, A. D. (2005). *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics (NCES 2005–464), figures 2–4, 3–5, 3–6, and 3–7. Data from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1973–2004 Long-Term Trend Mathematics Assessment*. Washington, DC: Government Printing Office.

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

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cogn. Psychol.* 41, 1–48. doi: 10.1006/cogp.1999.0735

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

Rodríguez-Santos, J. M., Calleja, M., García-Orza, J., Iza, M., and Damas, J. (2014). Quantity processing in deaf and hard of hearing children: evidence from symbolic and nonsymbolic comparison tasks. *Am. Ann. Deaf* 159, 34–44. doi: 10.1353/aad.2014.0015

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

Rudner, M., Orfanidou, E., Cardin, V., Capek, C., Woll, B., and Rönnberg, J. (2012). "Reading ability in adult deaf native signers is positively associated with their ability to judge the grammatically of their native sign language," in *Paper Presented at the Neurobiology of Language* (San Sebastian).

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. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372

Scott, G. D., Kams, C. M., Dow, M. W., Stevens, C., and Neville, H. J., (2014). Enhanced peripheral visual processing in congenitally deaf humans is supported by multiple brain regions, including primary auditory cortex. *Front. Hum. Neurosci.* 8, 1–9. doi: 10.3389/fnhum.2014.00177

Siegel, L. S., and Ryan, E. B. (1988). Development of grammatical-sensitivity, phonological, and short-term-memory skills in normally achieving and learning-disabled children. *Dev. Psychol.* 24, 28–37. doi: 10.1037/0012-1649.24.1.28

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

Swanwick, R., Oddy, A., and Roper, T. (2005). Mathematics and deaf children: an exploration of barriers to success. *Deafness Educ. Int.* 7, 1–21. doi: 10.1179/146431505790560446

Tavakoli, H. M. (2016). The relationship between accuracy of numerical magnitude comparisons and children's arithmetic ability: a study in iranian primary school children. *Eur. J. Psychol.* 12, 567–583. doi: 10.5964/ejop.v12i4.1175

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

Traxler, C. B. (2000). The Stanford achievement test, 9th edition: national norming and performance standards for deaf and hard-of-hearing students. *J. Deaf Stud. Deaf Educ.* 5, 337–348. doi: 10.1093/deafed/5.4.337

Tudusciuc, O., and Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 104, 14513–14518. doi: 10.1073/pnas.0705495104

Turconi, E., Jemel, B., Rossion, B., and Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cogn. Brain Res.* 21, 22–38. doi: 10.1016/j.cogbrainres.2004.05.003

Vederhus, L., and Krekling, S. (1996). Sex differences in visual spatial ability in 9-year-old children. *Intelligence* 23, 33–43. doi: 10.1016/S0160-2896(96)80004-3

Vitova, J., Zdrazilová, T., and Jezková, A. (2014). Successes of students with hearing impairment in math and reading with comprehension. *Proc. Soc. Behav. Sci.* 112, 725–729. doi: 10.1016/j.sbspro.2014.01.1223

von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., and Zhou, X. (2012). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychol. Sci.* 23, 320–330. doi: 10.1177/0956797611427168

Willingham, W., and Cole, N. (1997). *Gender and Fair Assessment*. Mahwah, NJ: Erlbaum.

Wu, C. M., Liu, T. C., Liao, P. J., Chen, C. K., Chang, B. L., and Lin, B. G. (2013). Academic achievements and classroom performance in mandarin-speaking prelingually deafened school children with cochlear implants. *Int. J. Pediatr. Otorhinolaryngol.* 77, 1474–1480. doi: 10.1016/j.ijporl.2013.06.012

Zarfaty, Y., Nunes, T., and Bryant, P. (2004). The performance of young deaf children in spatial and temporal number task. *J. Deaf Stud. Deaf Educ.* 9, 315–326. doi: 10.1093/deafed/enh034

Zhang, T., Chen, C., Chen, C., and Wei, W. (2020).Gender differences in the development of semantic and spatial processing of numbers. *Br. J. Dev. Psychol.* 38, 391–414. doi: 10.1111/bjdp.12329

Zhang, Y., Chen, C., Liu, H., Cui, J., and Zhou, X. (2016). Both non-symbolic and symbolic quantity processing are important for arithmetical computation but not for mathematical reasoning. *J. Cogn. Psychol.* 28, 1–18. doi: 10.1080/20445911.2016.1205074

Zhang, Y., Liu, T., Chen, C., and Zhou, X. (2019). Visual form perception supports approximate number system acuity and arithmetic fluency. *Learn. Individ. Differ.* 71, 1–12. doi: 10.1016/j.lindif.2019.02.008

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., and Dong, Q. (2007). Chinese kindergartners' automatic processing of numerical magnitude in Stroop-like tasks. *Mem. Cogn.* 35, 464–470. doi: 10.3758/BF03193286

Zhou, X., Wei, W., Zhang, Y., Cui, J., and Chen, C. (2015). Visual perception can account for the close relation between numerosity processing and computational fluency. *Front. Psychol.* 6:1364. doi: 10.3389/fpsyg.2015.01364

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

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 14 April 2021
doi: 10.3389/fpsyg.2021.629984

Arithmetic Errors in Financial Contexts in Parkinson's Disease

Hannah D. Loenneker[1], Sara Becker[2], Susanne Nussbaum[2], Hans-Christoph Nuerk[1†] and Inga Liepelt-Scarfone[2,3,4†]*

[1] Department of Psychology, Diagnostics and Cognitive Neuropsychology, Eberhard Karls Universität Tübingen, Tübingen, Germany, [2] Department of Neurodegeneration, Hertie Institute for Clinical Brain Research, Tübingen, Germany, [3] German Centre for Neurodegenerative Diseases (DZNE), Tübingen, Germany, [4] IB Hochschule für Gesundheit und Soziales, Stuttgart, Germany

Research on dyscalculia in neurodegenerative diseases is still scarce, despite high impact on patients' independence and activities of daily living function. Most studies address Alzheimer's Disease; however, patients with Parkinson's Disease (PD) also have a higher risk for cognitive impairment while the relation to arithmetic deficits in financial contexts has rarely been studied. Therefore, the current exploratory study investigates deficits in two simple arithmetic tasks in financial contexts administered within the Clinical Dementia Rating in a sample of 100 PD patients. Patients were classified as cognitively normal (PD-NC) or mildly impaired (PD-MCI) according to Level I consensus criteria, and assessed using a comprehensive neuropsychological test battery, neurological motor examination, and sociodemographic and clinical questionnaires. In total, 18% showed arithmetic deficits: they were predominately female, had longer disease duration, more impaired global cognition, but minor signs of depression compared to PD patients without arithmetic deficits. When correcting for clinical and sociodemographic confounders, greater impairments in attention and visuo-spatial/constructional domains predicted occurrence of arithmetic deficits. The type of deficit did not seem to be arbitrary but seemed to involve impaired place × value processing frequently. Our results argue for the importance of further systematic investigations of arithmetic deficits in PD with sensitive tests to confirm the results of our exploratory study that a specific subgroup of PD patients present themselves with dyscalculia.

Keywords: dyscalculia, financial management, neurodegeneration, MCI, elderly, gender differences, attention, visuo-spatial function

Edited by:
Júlia Beatriz Lopes-Silva, Federal University of Minas Gerais, Brazil

Reviewed by:
Francesca Burgio, University of Padua, Italy
Colin Joseph Hamilton, Northumbria University, United Kingdom

***Correspondence:**
Hannah D. Loenneker
hannah-dorothea.loenneker@uni-tuebingen.de

†These authors have contributed equally to this work

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

Received: 16 November 2020
Accepted: 12 March 2021
Published: 14 April 2021

Citation:
Loenneker HD, Becker S, Nussbaum S, Nuerk H-C and Liepelt-Scarfone I (2021) Arithmetic Errors in Financial Contexts in Parkinson's Disease. Front. Psychol. 12:629984. doi: 10.3389/fpsyg.2021.629984

INTRODUCTION

Arithmetic function deteriorates with age (Stemmler et al., 2013) and underlies elderly independent living skills (i.e., financial management; Finke et al., 2017). Despite this importance for activities of daily living (ADL), research on arithmetic deficits in elderly is scarce and primarily conducted in children (Kaufmann et al., 2013; Knops et al., 2017). Within the elderly population, neurodegeneration increases susceptibility to arithmetic deficits (e.g., Kalbe, 1999; Halpern et al., 2003; Arcara et al., 2019). While most research has been conducted in Alzheimer's Disease (AD), there are first hints that patients with other dementias, such as Parkinson's Disease (PD) dementia (PDD), also present with dyscalculia (Kalbe, 1999). Even though one might expect arithmetic deficits to have been assessed thoroughly given the extensive profiling of cognitive impairment in PD, research is scarce (e.g., Kalbe, 1999; Tamura et al., 2003). Unsystematic clinical observations show arithmetic errors in both advanced PDD (Kalbe, 1999) and non-demented PD patients (Tamura et al., 2003; Zamarian et al., 2006; Scarpina et al., 2017).

Previous research in PD operationalized arithmetic uniformly. However, developmental studies show dyscalculia resulting from distinct impairments in both specific numerical and domain-general cognitive functions (Jordan and Montani, 1997; Kaufmann et al., 2013). Domain-general cognitive functions required for arithmetic arise from different neuropsychological domains such as attention, working memory, language, executive or visuo-spatial function (Knops et al., 2017). Specific numerical prerequisites for arithmetic are heterogeneous, with magnitude being a core representation (e.g., Dehaene and Cohen, 1995). Furthermore, the application of calculation procedures is essential and impaired in AD (Mantovan et al., 1999). Another basis of multi-digit arithmetic is place-value integration (i.e., identification, activation, manipulation of digits within and between Arabic numbers; Nuerk et al., 2015). Analyzing specific errors then allows to infer underlying mechanisms (Nuerk et al., 2015): Erroneous magnitude processing shows as rounding errors within the correct decade (i.e., $3 \times 6 = 16$ not 18). Impaired calculation procedures can arise as operand (i.e., $3 \times 6 = 12$) or operation errors (i.e., 3×6 solved as $3+6$). Errors regarding decade value (i.e., $3 \times 6 = 28$), place ($3 \times 6 = 180$) or both ($3 \times 6 = 280$) stress an impaired place-value integration.

Financial capabilities are distinct and multidimensional skills (Marson, 2001, 2013), with arithmetic abilities such as number comprehension, principles, mental and written calculation being crucial prerequisites. Other cognitively mediated skills such as global cognitive function, short-term and working memory, (verbal) memory and learning, executive function, visuo-motor skills, decision making, financial conceptual knowledge, or instrumental ADL are associated with financial capabilities (Sherod et al., 2009; Lichtenberg et al., 2016; Arcara et al., 2019). Due to their complexity as higher order cognitive functions, financial capacities are prone to processes of aging and neurodegeneration (Willis, 1996; Marson et al., 2000). As arithmetic functioning is crucial, the current study explores PD-immanent arithmetic errors in financial contexts.

Arithmetic–specific cognitive deficits have not been well studied in PD yet possibly due to focusing on motor symptoms. Nowadays, PD is defined as a multisystem disorder affecting motor, autonomous, psychiatric and cognitive function (Postuma et al., 2015), with cognitive impairments often being confounded with motor symptoms (Das et al., 2016). PD-specific cognitive classifications continuously range from normal cognition (PD-NC) over mild cognitive impairment (PD-MCI) to PDD (Aarsland, 2016). These cognitive profiles are heterogeneous; old age, male gender, cortical cerebrospinal fluid (CSF) amyloid-beta 1–42 (Aβ42) pathology, depression and, most importantly presence of PD-MCI indicate susceptibility for PDD conversion (Irwin et al., 2012; Marras and Chaudhuri, 2016; Aarsland et al., 2017; Lin et al., 2018). Several factors, such as education, gender or work experience, have been shown to affect (numerical) healthy aging (Delazer et al., 2013; Lövdén et al., 2020) and might influence PD patients' arithmetic ability. However, the relationship between these specific profiles and arithmetic deficits has not been studied in PD, indicating the need to characterize patients making arithmetic errors to diagnose them timely for early interventions (Tucker-Drob, 2019).

Furthermore drawing inferences how arithmetic deficits affect PD patients from previous research focusing on AD is difficult, as similarity in clinical profiles is limited despite neuropathological overlaps (e.g., cholinergic impairments; Bohnen et al., 2003). Arithmetic deficits in AD (Rosselli et al., 1998) and early impairment of complex financial capacity in prodromal AD or MCI (Triebel et al., 2009; Marson, 2013, 2015) suggest a possible diagnostic value of arithmetic function for cognitive deterioration in PD, requiring clarification. Both the cognitive stage where arithmetic deficits first occur and the quality of impairments remain unknown. Arithmetic function in financial contexts is important for the autonomy and legal responsibilities of PD patients (Sherod et al., 2009; Marson, 2013; Arcara et al., 2019). Therefore, alteration in number cognition might also be arise in a prodromal stage of PDD, being investigated in this study. Therefore, it is crucial to phenotype arithmetic deficits by defining cognitive profiles of affected patients which can be achieved by addressing associations to other cognitive functions as suggested by different PD stages showing distinct cognitive profiles (Lopes et al., 2017).

The aim of the current study is (H1) to identify the frequency of arithmetic errors in financial contexts in PD-NC and PD-MCI patients and (H2) to profile characteristic patients committing these errors regarding sociodemographic, clinical, and cognitive measurements as compared to arithmetically unaffected patients. The last hypothesis addresses (H3) whether errors PD patients make can be attributed to specific categories of numerical processing to infer affected cognitive mechanisms. These hypotheses were investigated with data available from a longitudinal study focusing on the predictive value of CSF Aβ42 pathology in PD at the University Hospital in Tübingen. It includes sociodemographic and clinical assessments, as well as a neuropsychological test battery. Two arithmetic tasks in financial contexts administered within the Clinical Dementia Rating (Morris, 1993) were used to identify whether financial-arithmetic capabilities are a question relevant for PD research.

MATERIALS AND METHODS

Participants

Present data come from the longitudinal "Non-demented patients with Parkinson's disease with and without low Amyloid-beta 1–42 in cerebrospinal fluid" (ABC-PD longitudinal) study, focusing on the predictive value of Aβ42 pathology for cognitive worsening. The study was approved by the local ethics committee (686/2013BO1). Participants were recruited via the outpatient PD clinic or the ward at Tübingen University Hospital's neurology department. Patients received monetary compensation for travel expenses, and were assessed in the "on-state" with regular dopaminergic medication.

100 non-demented PD patients were selected by pre-screening neurological function confirming PD diagnosis following United Kingdom Brain-Bank criteria (Hughes et al., 1992). All patients received a lumbar puncture at least six weeks before the baseline visit and were between 50 and 85 years old. Patients were able to communicate well with the investigator, understand study requirements and give written informed consent. Diagnosis of PDD according to the Movement Disorder Society (MDS) Task Force criteria (Emre et al., 2007), other concomitant neurodegenerative diseases as well as substance abuse (except nicotine) led to participant exclusion.

Patients' CSF Aβ42 status was determined using commercially available ELISA kits (INNOTEST; Fujirebio Germany GmbH, Hannover, Germany, RRID: AB_2797385). Patients were divided into two equal sized (n = 50) groups: Aβ42+ (<600 pg/mL) and Aβ42– (≥600 pg/mL). Groups were matched according to age, gender and educational status. For the present analysis, PD-MCI was diagnosed according to the Level I MDS Task Force criteria (Litvan et al., 2012). All 100 patients were included in the analyses.

Material

Sociodemographic and Clinical Information

Demographics (age, gender, education years, disease duration, and age at PD onset) were acquired in an interview.

Arithmetic Function in Financial Contexts

Two standardized financial arithmetic tasks from the *Clinical Dementia Rating* interview (CDR; Morris, 1993; RRID:SCR_003678) were presented orally: *"How many 5 cent coins make up 1€?"* and *"How many 50 cent coins make up 15.50€?"* Answers in a verbal open response format were assessed based on correctness and errors where possible. Errors were ascribed to distinct categories (see **Table 1** for details and examples): Place-value integration errors (i.e., wrong decade value, wrong place, both), magnitude related errors (i.e., rounding in correct decade), procedural errors (i.e., operand error, wrong operation), or other errors (i.e., arbitrary errors, operands' repetition). Errors lacking information on participants' exact answers were categorized as NA.

TABLE 1 | Description of error categories with examples of patient answers.

Error category	Description	Patient answers	
		How many 5 cent coins make up 1€? = *20*	**How many 50 cent coins make up 15.50€? = *31***
Place-value	Wrong decade value	10	11
	Wrong place	2 or 200	
	Wrong place and decade value	100	
Magnitude	Rounding within correct decade		30 or 32
Procedural	Operand error (solution in multiplication table)	15	
	Wrong operation (division not multiplication)		7.75
	Wrong operation and place−value integration		76 (≈15.50 ÷ 2 ×10)
Others	Arbitrary	4	4 or 8 or 23 or 25 or 26
	Repetition of operand		15
NA	No information on exact error given		

Cognitive Function

The *Montreal Cognitive Assessment* (MoCA*;* Nasreddine et al., 2005) screened for global cognitive impairment (max. sum score 30 = normal cognitive performance). A MoCA score ≤ 26 indicated impaired global cognitive performance and assigned patients to the PD-MCI group. Cognitive function was additionally assessed with the German version of the *Repeatable Battery for the Assessment of Neuropsychological Status* (RBANS; Randolph et al., 1998). Scores on the twelve subtests were categorized into the domains attention, immediate and delayed memory, language, and visuo-spatial/constructional function (see **Supplementary Material B** for mapping of subtests to domains). Raw scores converted to age-group corrected z-scores, and composite domain and total scale scores according to the manual. The current analysis comprised RBANS domain and total scores.

Clinical Measurements

Parkinson's Disease motor symptoms were evaluated using the sum score of the MDS Unified Parkinson's Disease Rating Scale Part III (UPDRS III; Goetz et al., 2008) and Hoehn and Yahr staging (Hoehn and Yahr, 1967). The *UPDRS-III* rated motor symptoms on a scale ranging from 0 = normal to 4 = severe, with a maximum score of 132. The Hoehn and Yahr score ranging from one to four (1 = unilateral involvement; 4 = severe disability) additionally measured PD severity. Motor type was calculated from the UPDRS-III and item 12 from the former UPDRS-II version (Fahn et al., 1987) by means of the mean tremor score (postural, kinetic, or rest tremor) and the mean postural instability and gait disorder score (PIGD; falls, postural instability, freezing of gait). Patients were categorized as tremor-dominant in case of a ratio mean tremor score / mean PIGD score of 1.50 or higher or as PIGD dominant for ratios of 1.00 or lower, or as mixed for the remaining cases (Jankovic and Kapadia, 2001).

Anti-parkinsonian drug intake was expressed as *levodopa equivalent daily dose* (LEDD; Tomlinson et al., 2010). Patients' *depressive symptoms* during the last two weeks were rated with the Beck Depression Inventory (BDI-II; Hautzinger et al., 2006). *Health-related quality of life* was assessed with the single index score of the 39-item Parkinson's Disease Questionnaire (PDQ-39; Jenkinson et al., 1997). Items scored on a scale from 0 = never to 3 = often, with a successive transformation into weighted sum scores. The Functional Activities Questionnaire (FAQ; Pfeffer et al., 1982) was used to measure *activities of daily living function*. Patients rated their level of performance (0 = normal to 3 = dependent) on 10 ADLs subsumed as sum score.

Procedure

Testing took place in Tübingen University Hospital. Patients gave written informed consent and were assessed for eligibility based on the in- and exclusion criteria. Additionally to study assessments, most patients had appointments in the Parkinson's disease outpatient clinic before or after the study visit. Therefore, the order of clinical assessment varied between patients.

Data Analysis

Assessments were analyzed manually and data was managed using REDCap electronic data capture tools (Harris et al., 2009; RRID: SCR_003445). Statistical analyses were conducted with R version 4.0.3 (R Core Team, 2014; RRID: SCR_001905) and JASP version 0.13.1 (JASP Team, 2018; RRID: SCR_015823). Due to the small patient samples, Gaussian distribution of data was not assumed resulting in analyses using median, range, Mann-Whitney *U* tests, χ^2-tests, Brunner-Munzel tests (non-parametric trend test with $\hat{p}''$ = test statistic of stochastic equality; Brunner and Munzel, 2000), and binary logistic regressions. Confounders for the logistic regressions were chosen based on significantly differing variables between groups with and without arithmetic errors in financial contexts. Multicollinearity between predictors of the regression models was assessed based on a variance inflation factor (VIF) criterion above 10, not met by any predictor. For inferential statistics, an α-level of 0.05 was applied.

Analyses were conducted with the entire sample. Due to the over-representation of Aβ42+ patients, analyses were repeated with a subsample (N = 63) including all Aβ42- patients (n = 50) and a proportion of 20% Aβ42+ patients (n = 13). This reflects the estimated empirical distribution with a prevalence of AD pathology (Aβ42+, Tau, phosphorylated Tau) in approximately 30 to 40% of predominantly demented PD patients with non-demented PD patients falling considerably below this rate (Boller et al., 1980; Blennow and Hampel, 2003; Braak et al., 2005; Siderowf et al., 2010; Irwin et al., 2012). Unless indicated otherwise, outcomes did not differ between overall and representative sample (see **Supplementary Material**).

RESULTS

Frequency of Arithmetic Errors in Financial Contexts in PD (H1)

The total sample included 42% PD-MCI patients. Overall, 18% of PD patients (PD-NC and PD-MCI) showed arithmetic errors in at least one of the two financial items. PD-MCI patients showed 1 or 2 errors more frequently (26.2%) than PD-NC patients (12.1%), however, this difference did not reach significance, $\hat{p}''$ = 1.74, p = 0.09. For the Aβ42 groups, 16.0% of positive and 20.0% of negative patients showed arithmetic errors; this was not statistically significant $\chi^2(1) = 0.27$, p = 0.60.

The amount of errors differed marginally significantly between PD-NC (0 errors: 87.9%, 1 error: 8.6%, 2 errors: 3.4%) and PD-MCI (0 errors: 73.8%, 1 error: 4.8%, 2 errors: 21.4%), $\hat{p}'' = -1.92$, p = 0.059. The binary logistic regression correcting for the influence of gender, disease duration, and depression [$\chi^2(94) = 11.92$, p = 0.018, R^2_{McFadden} = 0.088, Area under the curve (AUC) = 0.695] revealed the amount of errors was the only significant predictor of cognitive status, with PD-MCI displaying more errors than PD-NC (p = 0.004). This model did not reach significance with the reduced representative sample (see **Supplementary Material A**). For analyses of RBANS subtests see **Supplementary Material B** (gender, depression, story memory predict arithmetic errors; differences between arithmetic groups: digit span, list learning, list recognition, picture naming, semantic fluency, line orientation).

Phenotyping Arithmetic Errors in PD (H2)

Patients with arithmetic errors differed from those without regarding gender (more females), disease duration (longer), depression (lower BDI-II scores), and global cognition (lower MoCA total and RBANS total scale scores, see **Table 2**). On average, females (0 error: 63.6%, 1 error: 12.1%, 2 errors: 24.2%) committed more errors than males (0 error: 91%, 1 error: 4.5%, 2 errors: 4.5%), $\hat{p}''$ = 3.01, p = 0.004. Effects were the same in the representative sample, except for groups not differing statistically regarding disease duration and depression (**Supplementary Material, Table A1**).

Results of the binary logistic regression indicated a significant association of gender, disease duration, depression, and all RBANS domain scores with the presence of arithmetic errors, $\chi^2(90) = 51.12$, $p < 0.001$, R^2_{McFadden} = 0.545, AUC = 0.941. Female gender, long disease duration, low depression scores, impaired attention and visuo-spatial/constructional deficits significantly predicted arithmetic errors (see **Table 3**). In the representative sample, only attention was a significant predictor for study group (see **Supplementary Material, Table A2**). In a second binary logistic regression model, confounding variables (gender, disease duration, depression) influenced presence of arithmetic errors $\chi^2(94) = 22.00$, $p < 0.001$, R^2_{McFadden} = 0.234, AUC = 0.822, while the FAQ score did not (see **Table 3**). This model was not stable in the representative sample (see **Supplementary Material A**).

Categorization of Errors (H3)

Arithmetic tasks differed regarding error categories in the entire patient sample. For the 5 cent task, place-value integration errors were most frequent (54.5%), followed by procedural (9.1%) and other errors (9.1%). In the 50 cent task, most errors could not be categorized (33.3%) or were magnitude-related (33.3%), followed by procedural (11.1%) and place-value integration errors (5.6%). Importantly, 27.3% (5 cent) and 16.7% (50 cent) of cases were NAs. The proportion of error categories did not differ between cognitive groups or by gender, but descriptively, more place-value integration errors occurred in the 5 cent task compared to the 50 cent task (see **Table 4**).

DISCUSSION

The current study aimed to identify the frequency of financial-arithmetic impairments in PD subgroups, as well as in relation to sociodemographic, clinical, and cognitive factors. Results demonstrate clinically relevant arithmetic errors in financial contexts. Identified risk factors were female gender, longer disease duration, greater severity of depressive symptoms, and more cognitive impairment. Place-value integration- and

TABLE 2 | Sociodemographic and clinical characterization of study patients.

	Total sample *N* = 100	Min. 1 arithmetic error *n* = 18	No arithmetic error *n* = 82	*p*
Age	65.49 (50.52–80.36)	66.80 (54.21–79.58)	64.60 (50.52–80.36)	0.15
Male *n* (%)	67.00 (67.00%)	6.00 (33.30%)	61.00 (74.40%)	<0.001*
Education years	13.00 (8.00–21.00)	12.00 (9.00–18.00)	13.00 (8.00–21.00)	0.16
Disease duration	4.17 (0.81–14.57)	6.51 (1.34–14.57)	3.98 (0.81–13.10)	0.02*
Age at onset	59.50 (38.94–77.53)	59.30 (39.79–74.78)	59.80 (38.94–77.53)	0.71
Aβ42+ status *n* (%)	50.00 (50.00%)	8.00 (44.40%)	42.00 (51.90%)	0.60
Motor type *n* (%)				0.87
PIGD	46.00 (46.00%)	9.00 (50.00%)	37.00 (45.10%)	
Mixed	10.00 (10.00%)	2.00 (11.10%)	8.00 (9.80%)	
Tremor dominant	44.00 (44.00%)	7.00 (38.90%)	37.00 (45.10%)	
UPDRS-III	25.00 (5.00–56.00)	23.00 (10.00–54.00)	25.00 (5.00–56.00)	0.89
Hoehn and Yahr score *n* (%)				0.25
1	3.00 (3.00%)	1.00 (5.60%)	2.00 (2.40%)	
2	79.00 (79.00%)	12.00 (66.70%)	67.00 (81.70%)	
3	18.00 (18.00%)	5.00 (27.80%)	13.00 (15.90%)	
LEDD	560 (100–2077.25)	627.50 (250–1353)	546.00 (100–2077.25)	0.42
BDI-II	6.00 (0–28.00)	3.50 (0–12.00)	6.00 (0–28.00)	0.03*
PDQ-39 summary index	2.15 (0.08–17.92)	2.76 (0.08–10.74)	2.13 (0.10–17.92)	0.66
FAQ	0 (0–20.00)	0 (0–15.00)	0 (0–20.00)	1.00
MoCA total score	26.00 (14.00–30.00)	24.50 (14.00–29.00)	26.00 (17.00–30.00)	0.02*
RBANS total scale score	90.00 (54.00–127.00)	82.00 (54.00–106.00)	92.50 (54.00–127.00)	0.002*

Group comparisons were conducted with Mann-Whitney U tests or χ^2 *where appropriate; median (range) or frequencies are given as measures of central tendency. *p < 0.05. Arithmetic errors are defined as: one or both arithmetic tasks not solved correctly. UPDRS-III = Unified Parkinson's Disease Rating Scale Part 3; LEDD = Levodopa equivalent daily dose; BDI = Beck Depression Inventory; PDQ-39 = Parkinson's Disease Questionnaire 39; MoCA = Montreal Cognitive Assessment; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status. Due to missing values, the overall sample was reduced to N = 99 for BDI-II, PDQ-39, and FAQ.*

TABLE 3 | Results of the binary logistic regressions predicting arithmetic errors in financial contexts.

						Wald Test			95%-CI		
	B	**SE**	**β**	**OR**	**z**	**Wald statistic**	**df**	***p***	**Lower bound**	**Upper bound**	**VIF**
1) Model including RBANS cognitive domain scores and covariates											
Intercept	16.58	6.23	−3.81	15860000	2.66	7.08	1	0.008*	4.37	26.79	
Gender	−3.29	1.11	−1.56	0.04	−2.97	8.79	1	0.003*	−5.47	−1.12	1.90
Disease duration	0.36	0.14	1.31	1.44	2.54	6.44	1	0.011*	0.08	0.64	1.51
BDI-II	−0.37	0.12	−2.31	0.69	−2.98	8.87	1	0.003*	−0.61	−0.13	1.92
Attention	−0.10	0.05	−1.72	0.91	−2.09	4.36	1	0.037*	−0.19	−0.01	3.12
Immediate memory	0.03	0.04	0.56	1.03	0.83	0.68	1	0.410	−0.04	0.10	2.88
Delayed memory	0.002	0.03	0.03	1.00	0.07	0.01	1	0.945	−0.06	0.07	1.99
Language	−0.034	0.05	−0.35	0.97	−0.63	0.40	1	0.529	−0.14	0.07	2.17
Visuo-spatial/constructional	−0.10	0.04	−1.35	0.91	−2.31	5.33	1	0.021*	−0.18	−0.01	1.94
2) Model including ADL function and covariates											
Intercept	−0.94	0.68	−2.07	0.39	−1.38	1.91	1	0.167	−2.26	0.39	
Gender	−1.72	0.63	−0.82	0.18	−2.76	7.59	1	0.006*	−2.95	−0.50	1.07
Disease duration	0.19	0.08	0.69	1.21	2.28	5.20	1	0.023*	0.03	0.36	1.14
BDI-II	−0.15	0.07	−0.96	0.86	−2.32	5.36	1	0.021*	−0.28	−0.02	1.23
FAQ	0.07	0.08	0.29	1.08	0.93	0.87	1	0.350	−0.08	0.28	1.31

**p < 0.05. B = estimated regression coefficient; SE = Standard error; β = standardized regression coefficient; OR = Odds Ratio; df = degrees of freedom; CI = confidence interval of the estimate; BDI = Beck Depression Inventory; FAQ = Functional activity questionnaire. Arithmetic errors level '1' was coded as class "min. 1 error" and gender level '1' was coded as "male."*

TABLE 4 | Proportion of error categories in relation to total errors as percentages per cognitive status and gender.

Error category	How many 5 cent coins make up 1€?						How many 50 cent coins make up 15.50€?					
	Cognitive status		*p*	Gender		*p*	Cognitive status		*p*	Gender		*p*
	PD-MCI	PD-NC		Male	Female		PD-MCI	PD-NC		Male	Female	
Place-value	55.6%	50.0%	1.00	66.7%	50.0%	1.00	9.1%	0%	1.00	16.7%	0%	0.25
Magnitude	0%	0%		0%	0%		27.3%	42.9%		16.7%	41.7%	
Procedural	11.1%	0%		0%	12.5%		9.1%	14.3%		0%	16.7%	
Others	11.1%	0%		33.3%	0%		45.5%	14.3%		50.0%	25.0%	
NA	22.2%	50.0%		0%	37.5%		9.1%	28.6%		16.7%	16.7%	

Group comparisons were conducted with χ^2 tests and frequencies are given as measures of central tendency. Both NAs (wrong answers without specification of the error committed) and the category others were excluded from the χ^2 tests.

magnitude-related errors were most frequent. Cognitive groups differed regarding the amount of errors in some analyses. Note the limitation that arithmetic errors increased for longer and cognitively more severe PD, but patients were not compared with controls.

Frequency of Arithmetic Errors in Financial Contexts in PD (H1)

The overall frequency of arithmetic errors of 18% in two simple tasks supports the need for further systematic investigation. When correcting for clinical confounders, the amount of errors was able to correctly predict cognitive status where higher errors were indicative of PD-MCI. However, some PD-NC patients also showed arithmetic errors, similar to previous AD studies finding dyscalculia in early stages (Parlato et al., 1992; Martin et al., 2003). Therefore, PD patients showing heterogeneous arithmetic impairments even in early stages and its association to specific cognitive profiles demands further examination. The model using the representative sample did not reach significance, which needs to be interpreted with care due to: a smaller sample (63 instead of 100 patients), an associated decrease in arithmetic errors, a greater tendency to ceiling effects, and a smaller amount of explained variance. Based on the current findings, it is impossible to infer on global financial capacities, as these are defined multidimensionally and exceed arithmetic function alone. However, showing a difference in arithmetic errors between PD-NC and PD-MCI indicates an association of PD disease severity and the likelihood of arithmetic errors in financial contexts.

Phenotyping Arithmetic Errors in Financial Contexts in PD (H2)

Profiles of PD patients with arithmetic errors were female, longer disease duration, less depression, and more cognitively impaired. Total MoCA and RBANS scale scores differed significantly between arithmetic groups, suggesting an association between arithmetic errors and cognition in PD. As PD-MCI and longer disease duration predict PDD (Aarsland et al., 2017), our data suggest that arithmetic errors occur within the frame of heterogeneous progressive cognitive deterioration.

When analyzing the proportion of errors (0,1,2), a systematic effect of gender (women made more errors) and cognitive status (PD-MCI patients made more errors) was observed, even after correcting for confounders. However, when error frequency was aggregated (1 or 2 errors in one category), when confounders were not considered, or when a smaller sample was used with a proportion of Aβ42+, differences between PD-NC and PD-MCI were only trends. We attribute this lack of significance when error categories are grouped to a statistical power issue due to sample sizes, tendency toward ceiling effects in the more representative sample, or decreased explained variance. While these results suggest more severe arithmetic errors in more cognitively impaired PD patients, they also advise for a future more systematic assessment.

The binary logistic regressions showed attention, visuo-spatial/constructional function, and story memory predicted arithmetic errors, suggesting degeneration in these domains at least partly causing arithmetic deficits. Current research in PD also discusses the importance of attention and visuo-spatial function for cognitive status und progression to PDD, introducing these domains as candidates for early biomarkers (Lopes et al., 2017; Becker et al., 2020). Associations between visuo-spatial/constructional functions and numerosity in healthy adults are in line with finding visuo-spatial/constructional functions to predict arithmetic errors in the current study (Lammertyn et al., 2002; Thompson et al., 2013). Furthermore, retrieving arithmetic facts requires an intact verbal memory (Dehaene and Cohen, 1997). Interestingly, females showed worse arithmetic performance than males. The gender differences in favor of men are in line with findings by Delazer et al. (2013) and Arcara et al. (2019). They explain advantages of elderly men in arithmetic and financial capabilities with employment in mathematics-related fields, higher level of education (paralleling generational effects) and mathematical interests. The arithmetic advantage for men is remarkable as gender effects usually dissociate from men showing stronger global cognitive decline but indicate differentiated visuo-spatial and verbal memory deficits in female PD patients (Fengler et al., 2016; Bakeberg et al., 2021). Therefore, the current association of visuo-spatial/constructional functions and story memory with arithmetic deficits is in line with finding more arithmetic deficits in female patients.

We also found that patients with arithmetic errors were, on average, less depressed than those without. This contradicts research on numeracy skills negatively affecting mental health in elderly (Fastame et al., 2019), and depression impairing cognitive performance in PD (Alzahrani and Venneri, 2015). As the difference in arithmetic does not seem to be PD-specific or directly related to cognition, the association with gender and education should be addressed to differentiate a coincidental finding from a systematic effect of depression on arithmetic in PD.

The binary logistic regression including ADL explained no more than a small amount of variance in arithmetic errors. The insignificant effect might be explained by patients in the current study being less heavily impacted on ADL, with ADL impairment occurring later in the process of transition from PD-MCI to PDD (Becker et al., 2020).

Presence of PD-Specific Error Categories (H3)

Most observed errors were place-value integration errors, followed by magnitude-related or procedural errors. Therefore, our data provide first evidence that not all numerical representations are impaired alike in PD. Future research should examine the validity of these results in a systematic methodological setup, with multi-digit and complex arithmetic tasks, enough items for a broader investigation of errors, and enough statistical power to identify PD-specific error categories and differences between cognitive statuses.

Limitations and Future Studies

The current exploratory study indicates the presence of particular arithmetic error types in financial contexts in PD, which – in some analyses – seem to associate with cognitive decline. These findings are novel in a hitherto neglected research field. However, this study has a couple of limitations, requiring consideration in follow-up studies.

First, the comparison with a healthy elderly group is missing to estimate the extent of impairments. However, finding arithmetic errors in both groups of PD-NC and PD-MCI with arithmetic errors to increase alongside progression of cognitive status indicates differing arithmetic impairments in discrete PD stages. Future studies should include healthy controls for comparisons between PD and the general population as well as more cognitively nuanced PD groups. Second, the methodological approach is not sufficient to provide generalizable inferences on arithmetic errors in financial contexts in PD with a rate of 18%. Yet, these errors are not negligible, but are absent in the diagnostic criteria for cognitive deficits in PD. Third, there were only two items and one type of arithmetic problem. Future studies should employ more systematic and broader assessments (e.g., different operations, different magnitudes, different place-value processing procedures, verbal and non-verbal tasks, symbolic and non-symbolic tasks) to obtain a more comprehensive overview on which errors are PD-specific and how pronounced they are for different numerical representations and processes. Fourth, this study is cross-sectional; however, longitudinal studies, informative for characterizing neurodegenerative processes when appropriately correcting for practice effects and selective attrition (Moody et al., 2017; Tucker-Drob, 2019), are missing. These designs can identify person-to-person heterogeneity in trajectories of lifespan cognitive developments being specific for the respective cognitive ability and interdependent for individuals (see Tucker-Drob, 2019). Heterogeneous arithmetic deficits found in AD (Girelli and Delazer, 2001) stress the need for differential investigations in PD.

Conclusion

In conclusion, while the current study reports first interesting data about arithmetic errors in financial contexts in PD, it is a mere starting point and inspiration to investigate such errors more systematically in the future. The current study suggests: (1) PD patients show arithmetic errors in financial contexts which seem to be more pronounced with cognitive impairment, (2) error type does not seem to be arbitrary but hints at a predominantly impaired place-value processing, and (3) apart from PD-MCI status there is heterogeneity within the groups and distinct attributes such as attention, visuo-spatial/constructional function, gender, or disease duration influence the likelihood of arithmetic errors, (4) The male advantage in arithmetic processing in PD contrasts men's larger global cognitive decline but follows female visuo-spatial and story memory disadvantages. In sum, we believe that these results suggest arithmetic performance in financial contexts to be a problem in PD-MCI, deserving future attention.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because public accessibility to the data has not been included in the ethics approval and participant's informed consent. Requests to access the datasets should be directed to HL, hannah-dorothea.loenneker@uni-tuebingen.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethik-Kommission an der Medizinischen Fakultät der Eberhard-Karls-Universität und am Universitätsklinikum Tübingen (686/2013BO1). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

IL-S and H-CN performed conceptualization. SB and SN performed data curation and investigation. HL performed formal analysis, visualization, and writing original draft. IL-S and HL performed funding acquisition. IL-S, HL, and H-CN performed methodology. SB, SN, and IL-S performed project administration. IL-S performed resources, supervision, and validation. HL, SB, H-CN, and IL-S performed writing review and editing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The "Non-demented patients with Parkinson's disease with and without low Amyloid-beta 1–42 in cerebrospinal fluid" study was funded by an unrestricted grant to IL-S from Janssen Research and Development, a division of Janssen Pharmaceutica N.V. The funding of this study was pre-competitive. HL received scholarships of the Landesgraduiertenfoerderung Baden-Wuerttemberg and of the Friedrich-Ebert Stiftung. H-CN is a member of the LEAD Graduate School and Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Wuerttemberg, and of the DFG research group "Modale und Amodale Kognition: Funktionen und Interaktionen" (FOR 2718, Teilprojekt: NU265/5-1).

ACKNOWLEDGMENTS

We thank the following persons supporting the design and conduction of the ABC-PD study: Daniela Berg, Walter Maetzler, Kathrin Brockmann, Maarten Timmers, Giacomo Salvadore, Wendy R. Galpern, and Johannes Streffer.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.629984/full#supplementary-material

REFERENCES

Aarsland, D. (2016). Cognitive impairment in Parkinson's disease and dementia with Lewy bodies. *Parkinsonism Relat. Disord.* 22, S144–S148. doi: 10.1016/J.PARKRELDIS.2015.09.034

Aarsland, D., Creese, B., Politis, M., Chaudhuri, K. R., Ffytche, D. H., Weintraub, D., et al. (2017). Cognitive decline in Parkinson disease. *Nat. Rev. Neurol.* 13, 217–231. doi: 10.1038/nrneurol.2017.27

Alzahrani, H., and Venneri, A. (2015). Cognitive and neuroanatomical correlates of neuropsychiatric symptoms in Parkinson's disease: a systematic review. *J. Neurol. Sci.* 356, 32–44. doi: 10.1016/j.jns.2015.06.037

Arcara, G., Burgio, F., Benavides-Varela, S., Toffano, R., Gindri, P., Tonini, E., et al. (2019). Numerical Activities of Daily Living – Financial (NADL-F): a tool for the assessment of financial capacities. *Neuropsychol. Rehabil.* 29, 1062–1084. doi: 10.1080/09602011.2017.1359188

Bakeberg, M. C., Gorecki, A. M., Kenna, J. E., Jefferson, A., Byrnes, M., Ghosh, S., et al. (2021). Differential effects of sex on longitudinal patterns of cognitive decline in Parkinson's disease. *J. Neurol.* 268, 1–10. doi: 10.1007/s00415-020-10367-8

Becker, S., Bäumer, A., Maetzler, W., Nussbaum, S., Tkaczynska, Z., Sulzer, P., et al. (2020). Association of cognitive activities of daily living (ADL) function and nonmotor burden in nondemented Parkinson's disease patients. *Neuropsychology* 34, 447-455. doi: 10.1037/neu0000627

Blennow, K., and Hampel, H. (2003). CSF markers for incipient Alzheimer's disease. *Lancet Neurol.* 2, 605–613. doi: 10.1016/S1474-4422(03)00530-1

Bohnen, N. I., Kaufer, D. I., Ivanco, L. S., Lopresti, B., Koeppe, R. A., Davis, J. G., et al. (2003). Cortical cholinergic function is more severely affected in Parkinsonian dementia than in Alzheimer disease. *Arch. Neurol.* 60, 1745–1748. doi: 10.1001/archneur.60.12.1745

Boller, F., Mizutani, T., Roessmann, U., and Gambetti, P. (1980). Parkinson disease, dementia, and alzheimer disease: clinicopathological correlations. *Ann. Neurol.* 7, 329–335. doi: 10.1002/ana.410070408

Braak, H., Rüb, U., Jansen Steur, E. N. H., Del Tredici, K., and De Vos, R. A. I. (2005). Cognitive status correlates with neuropathologic stage in Parkinson disease. *Neurology* 64, 1404–1410. doi: 10.1212/01.WNL.0000158422.41380.82

Brunner, E., and Munzel, U. (2000). The nonparametric behrens-fisher problem: asymptotic theory and a small-sample approximation. *Biom. J.* 42, 17–25.

Das, D., Biswas, A., Roy, A., Sauerbier, A., and Bhattacharyya, K. (2016). Cognitive impairment in idiopathic Parkinson's disease. *Neurol. India* 64, 419–427. doi: 10.4103/0028-3886.181533

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

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

Delazer, M., Kemmler, G., and Benke, T. (2013). Health numeracy and cognitive decline in advanced age. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 20, 639–659. doi: 10.1080/13825585.2012.750261

Emre, M., Aarsland, D., Brown, R., Burn, D. J., Duyckaerts, C., Mizuno, Y., et al. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Mov. Disord.* 22, 1689–1707. doi: 10.1002/mds.21507

Fahn, S., Eltron, R. L., and UPDRS Program Members (1987). "Unified Parkinson's disease rating scale," in *Recent Developments in Parkinson's Disease*, 2nd Edn, eds S. Fahn, C. D. Marsden, M. Goldstein, and D. B. Calne (Florham Park, NJ: Macmillan Healthcare Information), 153–163, 293–304.

Fastame, M. C., Manca, C., Penna, M. P., Lucangeli, D., and Hitchcott, P. K. (2019). Numeracy skills and self-reported mental health in people aging well. *Psychiatr. Q.* 90, 629–635. doi: 10.1007/s11126-019-09655-y

Fengler, S., Roeske, S., Heber, I., Reetz, K., Schulz, J. B., Riedel, O., et al. (2016). Verbal memory declines more in female patients with Parkinson's disease: the importance of gender-corrected normative data. *Psychol. Med.* 46, 2275–2286. doi: 10.1017/S0033291716000908

Finke, M. S., Howe, J. S., and Huston, S. J. (2017). Old age and the decline in financial literacy. *Manag. Sci.* 63, 213–230. doi: 10.1287/mnsc.2015.2293

Girelli, L., and Delazer, M. (2001). Numerical abilities in dementia. *Aphasiology* 15, 681–694. doi: 10.1080/02687040143000122

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., et al. (2008). Movement disorder society-sponsored revision of the Unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord.* 23, 2129–2170. doi: 10.1002/mds.22340

Halpern, C., McMillan, C., Moore, P., Dennis, K., and Grossman, M. (2003). Calculation impairment in neurodegenerative diseases. *J. Neurol. Sci.* 208, 31–38. doi: 10.1016/S0022-510X(02)00416-1

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). A metadata-driven methodology and workflow process for providing translational research informatics support. REDCap, Research electronic data capture. *J. Biomed. Inform.* 42, 377–381.

Hautzinger, M., Kühner, C., and Keller, F. (2006). *Das Beck Depressions Inventar II. Deutsche Bearbeitung und Handbuch zum BDI II*. Frankfurt: Harcourt Test Services.

Hoehn, M. M., and Yahr, M. D. (1967). Parkinsonism: onset, progression and mortality. *Neurology* 17, 427–442.

Hughes, A. J., Daniel, S. E., Kilford, L., and Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* 55, 181–184. doi: 10.1136/JNNP.55.3.181

Irwin, D. J., White, M. T., Toledo, J. B., Xie, S. X., Robinson, J. L., Van Deerlin, V., et al. (2012). Neuropathologic substrates of Parkinson disease dementia. *Ann. Neurol.* 72, 587–598. doi: 10.1002/ana.23659

Jankovic, J., and Kapadia, A. S. (2001). Functional decline in Parkinson disease. *Arch. Neurol.* 58, 1611–1615.

JASP Team (2018). *JASP (Version 0.13.1) [Computer software].*

Jenkinson, C., Fitzpatrick, R., Peto, V., Greenhall, R., and Hyman, N. (1997). The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. *Age Ageing* 26, 353–357. doi: 10.1093/ageing/26.5.353

Jordan, N. C., and Montani, T. O. (1997). Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *J. Learn. Disabil.* 30, 624–634. doi: 10.1177/002221949703000606

Kalbe, E. (1999). *Zahlenverarbeitung bei der Alzheimerschen Erkrankung und Anderen Demenzen*. Bielefeld: Universität Bielefeld.

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

Knops, A., Nuerk, H.-C., and Göbel, S. M. (2017). Domain-general factors influencing numerical and arithmetic processing. *J. Numer. Cogn.* 3, 112–132. doi: 10.5964/jnc.v3i2.159

Lammertyn, J., Fias, W., and Lauwereyns, J. (2002). Semantic influences on feature-based attention due to overlap of neural circuits. *Cortex* 38, 878–882. doi: 10.1016/s0010-9452(08)70061-3

Lichtenberg, P. A., Ficker, L., Rahman-Filipiak, A., Tatro, R., Farrell, C., Speir, J. J., et al. (2016). The lichtenberg financial decision screening scale (LFDSS): a new tool for assessing financial decision making and preventing financial exploitation. *J. Elder Abuse Neglect* 28, 134–151. doi: 10.1080/08946566.2016.1168333

Lin, S.-J., Baumeister, T. R., Garg, S., and McKeown, M. J. (2018). Cognitive profiles and Hub vulnerability in Parkinson's disease. *Front. Neurol.* 9:482. doi: 10.3389/fneur.2018.00482

Litvan, I., Goldman, J. G., Tröster, A. I., Schmand, B. A., Weintraub, D., Petersen, R. C., et al. (2012). Diagnostic criteria for mild cognitive impairment in Parkinson's disease: movement disorder society task force guidelines. *Mov. Disord.* 27, 349–356. doi: 10.1002/mds.24893

Lopes, R., Delmaire, C., Defebvre, L., Moonen, A. J., Duits, A. A., Hofman, P., et al. (2017). Cognitive phenotypes in parkinson's disease differ in terms of brain-network organization and connectivity. *Hum. Brain Mapp.* 38, 1604–1621. doi: 10.1002/hbm.23474

Lövdén, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U., and Tucker-Drob, E. M. (2020). Education and cognitive functioning across the life span. *Psychol. Sci. Public Interest* 21, 6–41. doi: 10.1177/1529100620920576

Mantovan, M. C., Delazer, M., Ermani, M., and Denes, G. (1999). The breakdown of calculation procedures in Alzheimer's disease. *Cortex* 35, 21–38. doi: 10.1016/S0010-9452(08)70783-4

Marras, C., and Chaudhuri, K. R. (2016). Nonmotor features of Parkinson's disease subtypes. *Mov. Disord.* 31, 1095–1102. doi: 10.1002/mds.26510

Marson, D. C. (2001). Loss of financial competency in dementia: conceptual and empirical approaches. *Aging Neuropsychol. Cogn.* 8, 164–181. doi: 10.1076/anec.8.3.164.827

Marson, D. C. (2013). Clinical and ethical aspects of financial capacity in dementia: a commentary. *Am. J. Geriatr. Psychiatry* 21, 382–390. doi: 10.1016/j.jagp.2013.01.033

Marson, D. C. (2015). Investigating functional impairment in preclinical Alzheimer's disease. *J. Prev. Alzheimers Dis.* 2, 4–6. doi: 10.14283/jpad.2015.44

Marson, D. C., Sawrie, S. M., Snyder, S., McInturff, B., Stalvey, T., Boothe, A., et al. (2000). Assessing financial capacity in patients with Alzheimer disease. *Arch. Neurol.* 57, 877–884. doi: 10.1001/archneur.57.6.877

Martin, R. C., Annis, S. M., Darling, L. Z., Wadley, V., Harrell, L., and Marson, D. C. (2003). Loss of Calculation abilities in patients with mild and moderate Alzheimer disease. *Arch. Neurol.* 60, 1585–1589. doi: 10.1001/archneur.60.11.1585

Moody, C. J., Mitchell, D., Kiser, G., Aarsland, D., Berg, D., Brayne, C., et al. (2017). Maximizing the potential of longitudinal cohorts for research in neurodegenerative diseases: a community perspective. *Front. Neurosci.* 11:467. doi: 10.3389/fnins.2017.00467

Morris, J. C. (1993). The clinical dementia rating (CDR): current version and scoring rules. *Neurology* 43, 2412–2412. doi: 10.1212/WNL.43.11.2412-a

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

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

Parlato, V., Lopez, O. L., Panisset, M., Iavarone, A., Grafman, J., and Boller, F. (1992). Mental calculation in mild Alzheimer's disease: a pilot study. *Int. J. Geriatr. Psychiatry* 7, 599–602. doi: 10.1002/gps.930070810

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., and Filos, S. (1982). Measurement of functional activities in older adults in the community. *J. Gerontol.* 37, 323–329. doi: 10.1093/geronj/37.3.323

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., et al. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* 30, 1591–1601. doi: 10.1002/mds.26424

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Randolph, C., Tierney, M. C., Mohr, E., and Chase, T. N. (1998). The repeatable battery for the assessment of neuropsychological status (RBANS): preliminary clinical validity. *J. Clin. Exp. Neuropsychol.* 20, 310–319. doi: 10.1076/jcen.20.3.310.823

Rosselli, M., Ardila, A., Arvizu, L., Kretzmer, T., Standish, V., and Liebermann, J. (1998). Arithmetical abilities in Alzheimer disease. *Int. J. Neurosci.* 96, 141–148. doi: 10.3109/00207459808986463

Scarpina, F., Mauro, A., D'Aniello, G. E., Albani, G., Castelnuovo, G., Ambiel, E., et al. (2017). Cognitive estimation in non-demented Parkinson's disease. *Arch. Clin. Neuropsychol.* 32, 381–390. doi: 10.1093/arclin/acx019

Sherod, M. G., Griffith, H. R., Copeland, J., Belue, K., Krzywanski, S., Zamrini, E. Y., et al. (2009). Neurocognitive predictors of financial capacity across the dementia spectrum: normal aging, mild cognitive impairment, and Alzheimer's disease. *J. Int. Neuropsychol. Soc. JINS* 15, 258–267. doi: 10.1017/S1355617709090365

Siderowf, A., Xie, S. X., Hurtig, H., Weintraub, D., Duda, J., Chen-Plotkin, A., et al. (2010). CSF amyloid β 1-42 predicts cognitive decline in Parkinson disease. *Neurology* 75, 1055–1061. doi: 10.1212/WNL.0b013e3181f39a78

Stemmler, M., Petermann, F., Daseking, M., Siebert, J., Schott, H., Lehfeld, H., et al. (2013). Diagnostik und verlauf von kognitiven fähigkeiten bei älteren Menschen. *Das Gesundheitswesen* 75, 761–767. doi: 10.1055/s-0033-1357164

Tamura, I., Kikuchi, S., Otsuki, M., Kitagawa, M., and Tashiro, K. (2003). Deficits of working memory during mental calculation in patients with Parkinson's disease. *J. Neurol. Sci.* 209, 19–23. doi: 10.1016/S0022-510X(02)00457-4

Thompson, J. M., Nuerk, H. C., Moeller, K., and Cohen Kadosh, R. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychol.* 144, 324–331. doi: 10.1016/j.actpsy.2013.05.009

Tomlinson, C. L., Stowe, R., Patel, S., Rick, C., Gray, R., and Clarke, C. E. (2010). Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov. Disord.* 25, 2649–2653. doi: 10.1002/mds.23429

Triebel, K. L., Martin, R., Griffith, H. R., Marceaux, J., Okonkwo, O. C., Harrell, L., et al. (2009). Declining financial capacity in mild cognitive impairment: a 1-year longitudinal study. *Neurology* 73, 928–934. doi: 10.1212/WNL.0b013e3181b87971

Tucker-Drob, E. M. (2019). Cognitive aging and dementia: a life-span perspective. *Annu. Rev. Dev. Psychol.* 1, 177–196. doi: 10.1146/annurev-devpsych-121318-085204

Willis, S. L. (1996). Everyday cognitive competence in elderly persons: conceptual issues and empirical findings. *Gerontologist* 36, 595–601. doi: 10.1093/geront/36.5.595

Zamarian, L., Visani, P., Delazer, M., Seppi, K., Mair, K. J., Diem, A., et al. (2006). Parkinson's disease and arithmetics: the role of executive functions. *J. Neurol. Sci.* 248, 124–130. doi: 10.1016/J.JNS.2006.05.037

Conflict of Interest: This study received funding from Janssen Research and Development, a division of Janssen Pharmaceutica N.V. The study is purely precompetitive. The funder was not involved in the study design and data collection. Additionally, the funder was not involved in the current analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 May 2021
doi: 10.3389/fpsyg.2021.633077

From Non-symbolic to Symbolic Proportions and Back: A Cuisenaire Rod Proportional Reasoning Intervention Enhances Continuous Proportional Reasoning Skills

Roberto A. Abreu-Mendoza[1], Linsah Coulanges[1†], Kendell Ali[2], Arthur B. Powell[3] and Miriam Rosenberg-Lee[1,4]*

[1] Department of Psychology, Rutgers University, Newark, NJ, United States, [2] Graduate School of Education, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States, [3] Department of Urban Education, Rutgers University, Newark, NJ, United States, [4] Center for Molecular and Behavioral Neuroscience, Rutgers, The State University of New Jersey, Newark, NJ, United States

OPEN ACCESS

Edited by:
Júlia Beatriz Lopes-Silva,
Federal University of Minas Gerais,
Brazil

Reviewed by:
Edward M. Hubbard,
University of Wisconsin-Madison,
United States
Larissa Salvador,
Universidade Federal de Minas
Gerais, Brazil

***Correspondence:**
Roberto A. Abreu-Mendoza
roberto.abreu@rutgers.edu

†Present address:
Linsah Coulanges,
Department of Psychology, University
of Pittsburgh, Pittsburgh, PA,
United States

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

Received: 24 November 2020
Accepted: 23 April 2021
Published: 21 May 2021

Citation:
Abreu-Mendoza RA, Coulanges L,
Ali K, Powell AB and
Rosenberg-Lee M (2021) From
Non-symbolic to Symbolic
Proportions and Back: A Cuisenaire
Rod Proportional Reasoning
Intervention Enhances Continuous
Proportional Reasoning Skills.
Front. Psychol. 12:633077.
doi: 10.3389/fpsyg.2021.633077

The persistent educational challenges that fractions pose call for developing novel instructional methods to better prepare students for fraction learning. Here, we examined the effects of a 24-session, Cuisenaire rod intervention on a building block for symbolic fraction knowledge, continuous and discrete non-symbolic proportional reasoning, in children who have yet to receive fraction instruction. Participants were 34 second-graders who attended the intervention (intervention group) and 15 children who did not participate in any sessions (control group). As attendance at the intervention sessions was irregular (median = 15.6 sessions, range = 1–24), we specifically examined the effect of the number of sessions completed on their non-symbolic proportional reasoning. Our results showed that children who attended a larger number of sessions increased their ability to compare non-symbolic continuous proportions. However, contrary to our expectations, they also decreased their ability to compare misleading discretized proportions. In contrast, children in the Control group did not show any change in their performance. These results provide further evidence on the malleability of non-symbolic continuous proportional reasoning and highlight the rigidity of counting knowledge interference on discrete proportional reasoning.

Keywords: proportional reasoning, intervention, non-symbolic processing, symbolic processing, inhibitory control

INTRODUCTION

Learning fractions is an arduous and protracted process for students. In the United States, fraction instruction typically starts in third grade with fraction expressions; then, in fourth grade, children are first introduced to arithmetic operations with fractions (Common Core State Standards Initiative, 2020b). However, even after four years of instruction, less than a third of eighth-graders (∼30%) show an understanding of fraction addition (Carpenter et al., 1980; Lortie-Forgues et al., 2015). Regrettably, fraction arithmetic is just one example of students' persistent difficulties with fractions (Siegler and Lortie-Forgues, 2017; van Hoof et al., 2018). Recent efforts from researchers and educators to develop novel methods involving non-symbolic representations to teach fractions are beginning to bear fruit. In the current study, we examined the effects of an intervention using Cuisenaire rods to improve non-symbolic proportional reasoning, a building block for symbolic fraction knowledge.

Part-Whole and Alternative Models of Fractions

Traditionally, fractions are represented using part-whole models (e.g., pie charts). However, these representations might impede understanding of fundamental fraction properties, such as ratio, by promoting whole-number strategies, like counting (Plummer et al., 2017). In contrast, interventions that use non-symbolic continuous models, like number lines, provide a shared representation for whole and rational numbers, take advantage of spatial-numeric associations and capture the continuous property of fractions (Hamdan and Gunderson, 2017). These interventions also leverage children's early proficiency at comparing and matching continuous proportional information (Boyer et al., 2008; Boyer and Levine, 2015; Hurst and Cordes, 2018). Indeed, one of the most promising methods to improve fraction skills is intensive training involving mapping non-symbolic continuous representations of proportions with fractions (Fazio et al., 2016; Braithwaite and Siegler, 2020; Soni and Okamoto, 2020; Wortha et al., 2020).

Emerging evidence from individual difference studies and experimental research also supports the link between non-symbolic continuous representations of proportions and fraction skills (Matthews et al., 2016; Bhatia et al., 2020; Kalra et al., 2020). For instance, college students who are more precise in their judgments of non-symbolic ratios are also better at comparing symbolic fractions (Matthews et al., 2016). Moreover, matching non-symbolic continuous representations of proportions to symbolic fractions is modulated by the distance effect (e.g., lower performance in comparing smaller ratios than larger ratios), suggesting that both formats activate the same mental proportional magnitude representations (Bhatia et al., 2020). Overall, these findings indicate that students might use their non-symbolic continuous proportional reasoning skills as a scaffold for symbolic fraction knowledge. Yet, little is known about the malleability of non-symbolic proportional reasoning through training.

Training Non-symbolic Proportional Reasoning

Continuous Non-symbolic Proportions

To date, only two studies have reported changes in non-symbolic continuous proportional reasoning following training. These studies employed individual, computerized interventions with carefully matched control conditions (Gouet et al., 2020; Wortha et al., 2020). In Gouet et al. (2020), nine-year-old children went through one of two non-symbolic continuous proportional interventions or an absolute magnitude control condition. In both non-symbolic interventions, children used a number line to estimate proportional continuous quantities, either the red area of a two-color rectangle or the size of a yellow circle relative to a blue circle. After the five-day intervention, children from both interventions improved their non-symbolic proportional skills, while children in the control group, who only practiced absolute magnitude comparison skills, did not. Crucially, the intervention also had a positive effect on children's symbolic fraction arithmetic and comparison skills.

In the second study, Wortha et al. (2020) examined the effects of fraction intervention on adults' reaction times and brain activation while performing three tasks: a cross-format proportional matching task, a number line comparison task, and a fraction comparison task. Their results showed that after estimating fraction magnitudes using number lines for five days, participants became more precise in matching number lines to fractions and comparing number lines after the intervention. However, they showed no gains in their symbolic fraction comparison skills. The brain imaging results showed the opposite pattern: there were no changes in the activation during the matching and number line comparison tasks, but during the symbolic fraction comparison task, activation increased in a set of frontoparietal regions implicated in math cognition, including the bilateral intra-parietal sulcus and the middle and the inferior gyrus. Together, these studies suggest that non-symbolic continuous proportional reasoning can be improved by exclusively training non-symbolic skills or by matching symbolic and non-symbolic proportions.

These studies employed strictly controlled computerized interventions, which lack ecological validity, and their implementation in traditional classrooms might be technologically challenging. In contrast, for the current study, we implemented a proportional reasoning intervention in classrooms and with inexpensive physical manipulatives. Using well-known educational materials, Cuisenaire rods, children transitioned from comparing the relative lengths of pairs of rods to expressing those comparisons symbolically (**Figures 1A,B**). Thus, our first aim was to examine the possible positive transfer effects of this intervention on the ability to compare proportions presented in another non-symbolic format (i.e., annulus-shaped figures, **Figure 1C**) in second-grade children who have yet to receive formal fraction instruction.

Discrete Non-symbolic Proportions

The ease with which children can represent non-symbolic continuous proportions contrasts with the great difficulty they encounter when comparing discrete proportions (Jeong et al., 2007; Boyer et al., 2008; Hurst and Cordes, 2018). For instance, while even four-year-old children can successfully compare non-symbolic continuous proportions, ten-year-old children still struggle with discrete ones, particularly when the discrete information contradicts proportional information (Jeong et al., 2007). Children's tendency to exert whole-number strategies to proportional reasoning tasks is consistent with similar phenomena seen in symbolic fraction comparison, an effect termed whole number bias (Ni and Zhou, 2005). Given the persistent developmental challenge entailed by discrete proportional reasoning, an outstanding question is whether

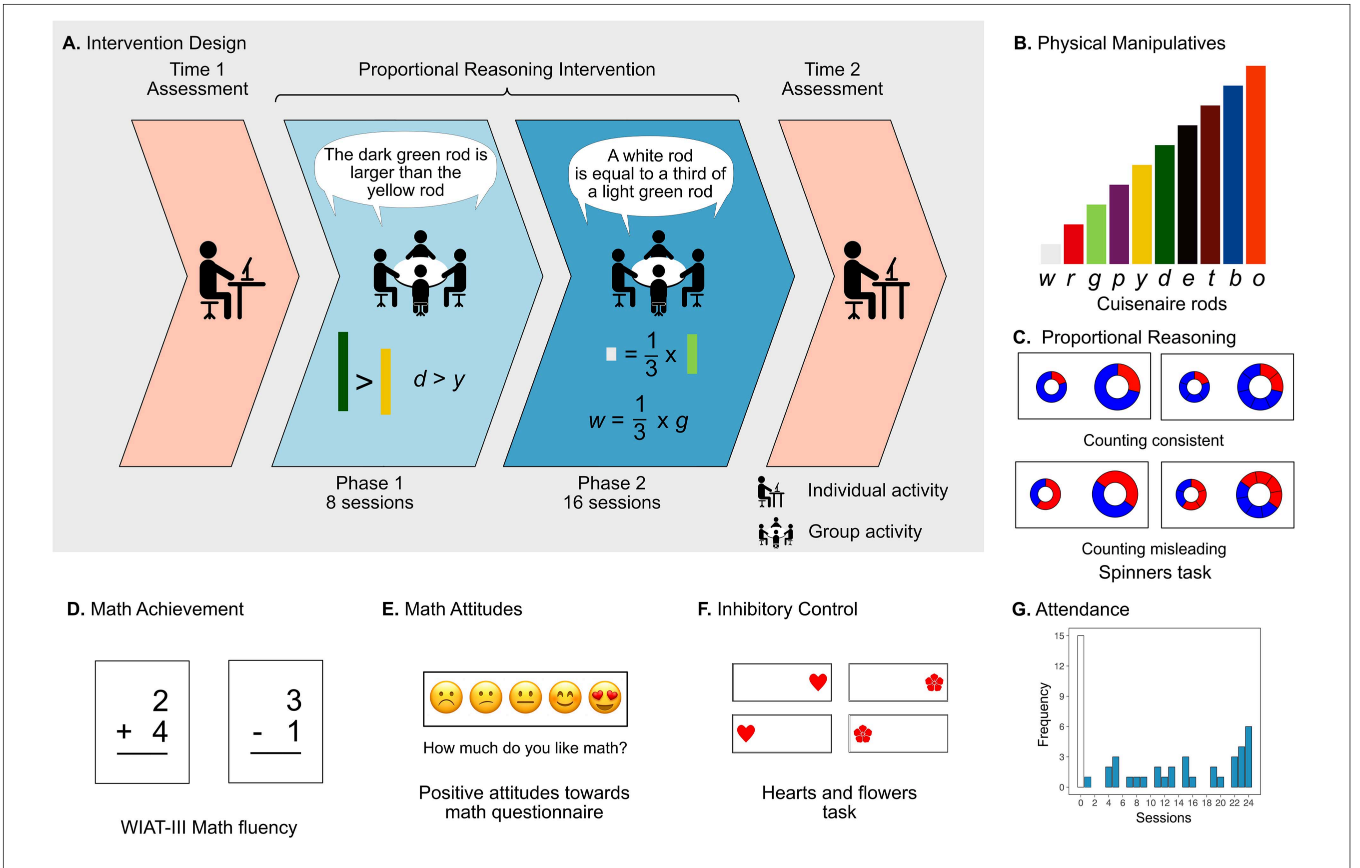


FIGURE 1 | Study design. **(A)** The study consisted of three stages: the time 1 assessment (late September 2018), 24 sessions of a group-based proportional reasoning intervention, and (3) a time 2 assessment (March 2019), which repeated the time 1 assessment. The intervention was further divided into two phases: Phase 1 (8 sessions), in which children became familiarized with the materials and manipulated absolute magnitudes, physically, orally and symbolically; Phase 2 (16 sessions), in which children represented and compared non-symbolic and symbolic proportions. **(B)** During both phases, children initially worked with the physical rods and then they used letters to refer to the rods (w = white, r = red, g = green, p = purple, y = yellow, d = dark green, e = ebony, t = tan, b = blue, o = orange). The time 1 and time 2 assessment comprised of **(C)** the spinners task, the main outcome task, which comprised continuous (left pairs spinners) and discrete (right pairs of spinners) formats **(D)** the arithmetic fluency subtests from the Wechsler Individual Achievement Test (WIAT)-III, **(E)** the positive attitudes toward math questionnaire, and **(F)** the hearts and flowers task. **(G)** The histogram shows the number of children who attended a given number of sessions, from zero sessions (control group) to 24 sessions (complete intervention).

interventions that improve continuous proportional reasoning influence discrete proportional reasoning skills. The second aim of this study is to shed light on this question.

A small set of studies have investigated how non-symbolic proportional comparison skills relate across formats (Mock et al., 2018; Park et al., 2020). Recently, Park et al. (2020) examined non-symbolic proportional comparison skills in preschoolers, second graders, fifth graders, and adults across continuous (circles, lines, and blob areas) and discrete (collections of circles) non-symbolic formats. The authors also evaluated absolute magnitude comparison skills across these four formats. They report that proportional skills in one format were better predicted by proportional skills in another format than absolute magnitude comparison skills of the same format. For instance, comparing the ratio between pairs of circles was better predicted by comparing the ratio between pairs of lines than comparing the absolute magnitude of circles. These results suggest that individuals use the same proportional comparison capacity regardless of the format in which proportions are presented. Consistent with these results, Mock et al. (2018) found in adults, overlapping brain regions for the processing of proportions presented as non-symbolic continuous and discrete representations, fractions, and decimals, as do other rational numbers (Rosenberg-Lee, 2021). These regions were the superior parietal lobule, the inferior, middle, and superior occipital gyri. Together, these studies suggest that improvements in one format should also be reflected in other formats due to proportional magnitudes being processed in an amodal manner. However, this conclusion is difficult to reconcile with the persistent decrements in performance found in non-symbolic discrete formats (Jeong et al., 2007; Begolli et al., 2020). An alternative line of research (Boyer and Levine, 2015; Hurst and Cordes, 2018; Abreu-Mendoza et al., 2020) suggests priming continuous proportional reasoning immediately before discrete stimuli mitigates the challenges of discrete proportional reasoning. The current study aimed to examine whether an intervention focused on non-symbolic continuous skills positively shapes children's non-symbolic discrete proportional reasoning skills.

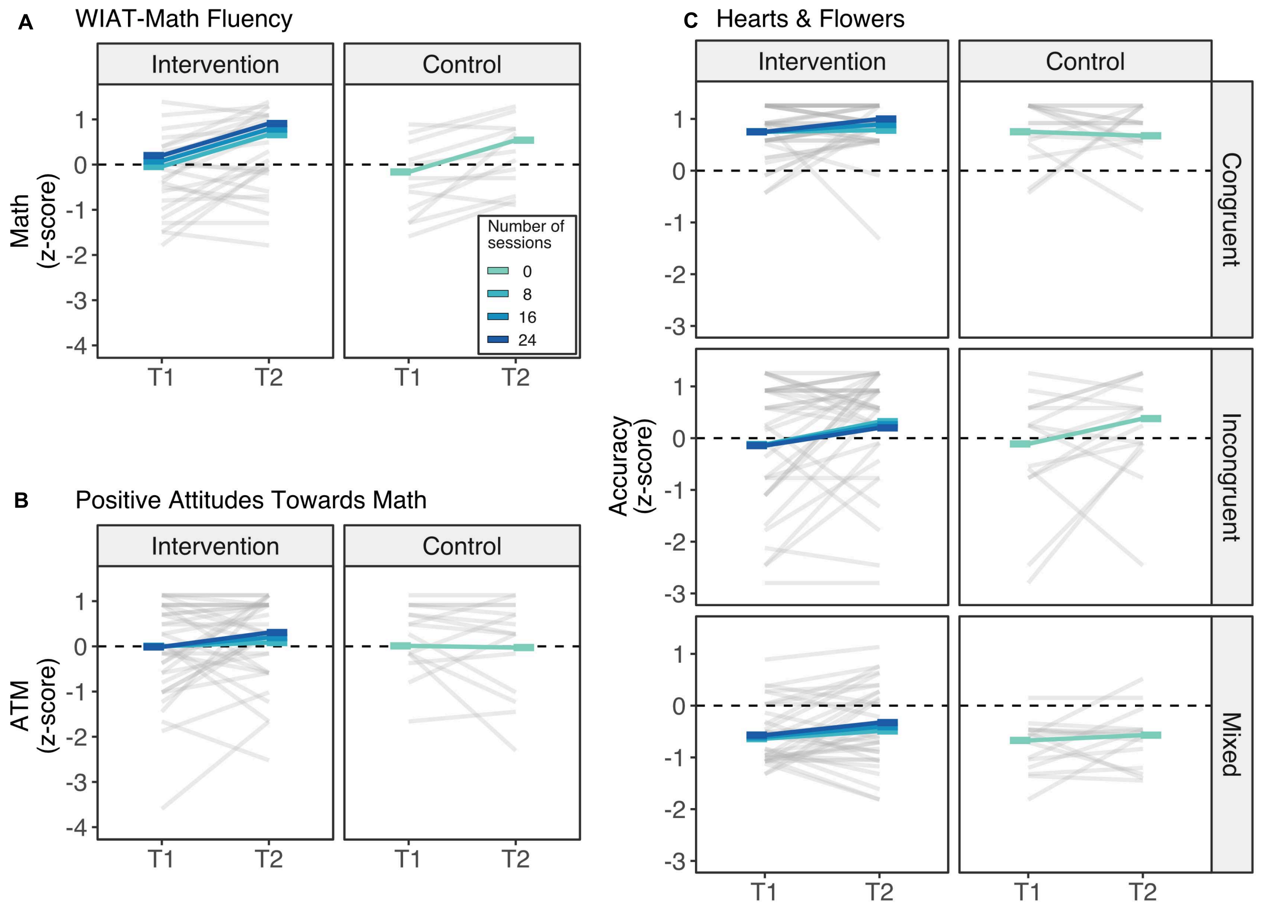


FIGURE 2 | Time 1 (T1) to time 2 (T2) changes in z-scores of the **(A)** WIAT-III Math Fluency composite scores, **(B)** Positive attitudes toward math scores, and **(C)** Hearts and Flowers' accuracy. Over the five months between time 1 and time 2 and regardless of the number of completed sessions of the intervention, children improved their math skills above and beyond what was expected during this period **(A)**. While children's attitudes toward math remained stable from time 1 to time 2 **(B)**, children also improved their performance in the incongruent block of Hearts and Flowers tasks **(C)**. Rods represents the estimated marginal means across different number of sessions of attendance from 0 sessions (control group) to 24 sessions (darkest blue).

Domain-Specific and Domain-General Predictors of Intervention Effects

The current intervention design also allowed us to examine the predictor effects of mathematical achievement and attitudes, as well as cognitive skills that previous intervention studies have shown play a critical role in fraction learning.

Math Abilities

Only a few studies have examined the moderating role of student's general math knowledge on fraction learning. In one study, Fuchs et al. (2013) showed that children's initial scores in the National Assessment of Educational Progress assessment did not influence children's gains from a 12-week fraction intervention. Conversely, Reinhold et al. (2020) found that prior math achievement, measured by the type of school (high and low achieving institutions), moderated the effects of different types of fraction interventions. High-achieving children showed larger gains from a new fraction curriculum than a traditional one, regardless of whether it was presented as either a book or an e-book. However, low-achieving children only benefited from the new curriculum when it was offered as an e-book. Longitudinal studies of fraction learning have shown that initial general math performance is a predictor for later conceptual and procedural fraction knowledge (Jordan et al., 2013). Together, these studies suggest the potential for a positive relationship between general math knowledge and fraction instruction; however, conclusive evidence is still emerging.

Attitudes Toward Math

Holding more positive attitudes toward mathematics is positively related to math performance (Chen et al., 2018; Dowker et al., 2019). Yet, little is known about the specific relations of math attitudes to non-symbolic and symbolic proportional reasoning. Recently, Sidney et al. (2019) found that children and adults had more negative attitudes toward fractions than whole numbers. Further, while children's attitudes toward whole-numbers and fractions were equally related to general math performance, adults' attitudes toward whole-numbers were more strongly associated with math than attitudes toward fractions. In the current study, we investigated whether positive attitudes toward math, in general, are predictive of learning gains in non-symbolic proportional reasoning.

Executive Functions and Inhibitory Control

Inhibitory control plays a critical role in acquiring information that contradicts previously learned knowledge both in science and math (Brookman-Byrne et al., 2018). Specific to fraction learning, inhibitory control may help students override automatic whole-number representations and hone in proportional magnitudes (Vosniadou et al., 2018). Studies of individual-differences finds that students with higher inhibition capacity are better at comparing misleading non-symbolic discrete proportions (Abreu-Mendoza et al., 2020), misleading fractions (Gomez et al., 2015; Avgerinou and Tolmie, 2019), and misleading decimals (Avgerinou and Tolmie, 2019; Coulanges et al., 2021). Here, we aimed to extend these findings by examining the predictive role of inhibitory control in learning gains in non-symbolic proportional reasoning, specifically when there is a need to disregard misleading discrete information.

Although these individual-difference studies suggest a positive relationship between learning gains and inhibitory control, a previous study of the moderator effect of working memory, another canonical executive function (Diamond, 2013), on fraction interventions alludes to a more nuanced relationship (Fuchs et al., 2014). In their study, the contribution of working memory varied depending on the type of intervention. While participants with low working memory benefitted most from a conceptually rich fraction intervention, participants with high working memory levels showed the largest gains when the intervention focused on fraction arithmetic fluency. Overall, these findings indicate that inhibitory control may play a key role in improving non-symbolic proportional reasoning, but its effects could depend on the kind of instruction.

The Current Study

The aims of this study were threefold: (1) provide further evidence on the malleability of non-symbolic continuous proportional reasoning in the context of a classroom-based, physical manipulatives intervention; (2) investigate whether an intervention that targets continuous representations of fractions leads to improvements in discrete proportional reasoning; (3) examine possible academic, attitudinal, and cognitive predictors of children's improvement in non-symbolic proportional reasoning. To achieve these study aims, second graders participated in a 24-session intervention program that introduced fractions as multiplicative comparisons between two continuous quantities. Specifically, students measured the length of rods of different sizes and learned to communicate in oral and written forms the relationship between the rod lengths.

Consistent with previous findings, we predicted that children's non-symbolic continuous proportional reasoning would increase following the intervention. We further hypothesized that discrete comparison skills, particularly in contexts where discrete information is misleading, would also improve. Based on the finding that inhibitory control relates to misleading discrete proportional reasoning, we tested the hypothesis that children with high inhibition skills would show larger learning gains in discrete misleading trials. These two hypotheses were pre-registered on a pre-registration poster submission (Mathematical Cognition and Learning Society, 2019). Finally, we predict that children with strong initial math skills and more positive attitudes toward math will show larger proportional reasoning gains than those with low math skills and less favorable attitudes.

MATERIALS AND METHODS

Participants and Intervention Phases

Fifty-seven students from three second-grade classrooms at a public school in Newark, NJ, were invited to participate in this study. Of the 52 children whose parents consented for them to participate in the study, 50 participants completed the two assessment sessions. Out of these students, 35 children were part of an after-school enrichment program that comprised the original intervention group. Among these 35 children, the final sample excluded one participant because they did not have the minimum number of usable trials in the key outcome measure (see Section "Proportional Reasoning"). Thus, the final sample of the intervention group consisted of 34 children. However, as attendance at the intervention sessions was irregular (median = 15.6 sessions, range = 1–24, **Figure 1G**), we used the number of sessions completed as an independent variable for all our analyses. Children who did not participate in the after-school program (n = 15) comprised the original control group, and they were coded as having 0 sessions in any analyses which used attendance as a continuous variable. All children's parents gave written informed consent, and the children gave oral assent for their participation. The Rutgers University Institutional Review Board approved the research protocol. The pre-training data from this sample is reported elsewhere (Abreu-Mendoza et al., 2020).

We performed a sensitivity power analysis using the final sample size (n = 49) as a reference for the planned correlations. These analyses indicated that our sample size enables detecting moderate to large correlations (Pearson's $r > 0.38$) using an alpha = 0.05 and power = 0.80. Using the same alpha and power values, in the intervention group, we can detect medium to large correlations (Pearson's $r > 0.44$) while in the control group only large effects (Pearson's $r > 0.62$).

Study Overview

The study consisted of three stages: (1) a pre-intervention assessment (time 1, late September 2018), which consisted of four activities, administered in the following order: Math fluency subtests (math achievement), Spinners task (proportional reasoning), Hearts and Flowers (H&F, inhibitory control) task, and Positive Attitudes toward Math questionnaire; (2) 24 sessions of a group-based proportional reasoning intervention; and (3) a post-intervention assessment (time 2, March 2019), in which participants performed the same activities as the pre-intervention assessment, in the same order (**Figure 1**). Children in the control group only completed stages 1 and 3.

The two computerized tasks, the Spinners and H&F tasks, were presented using PsychoPy2 Experiment Builder Version 1.90.3 (Pierce, 2007). Children were evaluated individually by trained experimenters in quiet corners of a large room at the children's school (maximum of 4 children at a time). Experimenters were blind to the group assignment of participants. Each assessment session lasted approximately 25 min.

Academic, Cognitive, and Attitude Assessments

Mathematical Achievement

As fractions are not typically taught in second grade in the United States (Common Core State Standards Initiative, 2020b), to evaluate children's mathematical achievement, we concentrated on skills appropriate for children's academic stage (i.e., arithmetic skills). Thus, we used the Math Fluency–Addition and Math Fluency–Subtraction subtests from the Wechsler Individual Achievement Test–Third Edition (WIAT-III; Wechsler, 2009). In each subtest, children answered as many arithmetic (first addition, then subtraction) problems as they could in 1 min. Combining the grade-normed scores of each subtest provides an age-appropriate measure of children's mathematical achievement.

Attitudes Toward Mathematics

To evaluate children's attitudes toward math, we adapted the 5-point Likert-type Positive Attitude toward Math (PAM) questionnaire (Chen et al., 2018). To make it appropriate for children, we used emojis to help connote the response options. This questionnaire was comprised of six items that evaluate children's attitudes toward math and six items that evaluate their general attitude toward academics (e.g., science, reading, computers, and technology). For this study, our variable of interest was the average of the first six questions relating to math attitudes.

Inhibitory Control

Children's inhibitory control was assessed with the Hearts and Flowers task (hereafter H&F task; Davidson et al., 2006; Brocki and Tillman, 2014; Wright and Diamond, 2014). This computerized task consisted of three blocks presented in the following fixed order: congruent, incongruent, and mixed. The experimenter read aloud the on-screen instructions to the children. In the congruent block, children were instructed to press the key on the same side as where the target (hearts) appeared, using the keys "z" for the left side and "m" for the right side. In the incongruent block, children were instructed to press the key on the opposite side from where the target (flowers) appeared. In the mixed block, children saw interspersed hearts and flowers and were asked to respond according to the previously learned rules. At the beginning of the congruent and incongruent blocks, there were 2 example trials. The corresponding figure (heart or flower) appeared first on the right and then on the left. The target images remained on the screen until the children pressed the correct key. The first two blocks comprised 12 trials each, which randomly presented the corresponding figure on each side six times. The third (mixed) block contained 33 trials, and the first trial of this block was always a heart presented on the right side. Subsequent trials randomly presented each figure 16 times, eight times on each side. We considered this last block as the measure of inhibitory control because prior research found that performance in the mixed block was strongly correlated with a latent variable of inhibition (r = 0.71), whereas performance in the congruent block and performance in the incongruent block were negative (r = −0.03) and weakly associated (r = 0.17), respectively (Brocki and Tillman, 2014).

Following Wright and Diamond (2014), when computing accuracy, we excluded anticipatory responses (reaction times [RT] shorter than 250 ms) and outlier responses (RTs at least 3 standard deviations above the individual's mean). After applying these criteria, among the 49 children, we analyzed 4397 (96.61%) of 5586 trials.

Outcome Task

Proportional Reasoning

To measure children's learning gains in proportional reasoning, we used a computerized version of the Spinners task (Jeong et al., 2007). In this task, children saw two spinners and had to indicate which of them has a proportionally larger red area.

The 12 proportions used by Jeong et al. (2007) were presented in three different format blocks for a total of 36 experimental trials. In the continuous format, each spinner had only two continuous sections, one red and one blue. In the discrete adjacent format, the two continuous parts were segmented into discrete but adjacent sections of red and blue sections. In the discrete mixed format, the red and blue segments were interspersed. In the discrete blocks, the number of segments was manipulated so that in half of the trials, the spinner with the larger number of red pieces was also the one with the proportionally larger red area (counting consistent trials). In contrast, in the other half, the spinner with the fewer red pieces was the one with the proportionally larger red area (counting misleading trials). Although "counting" information could not be meaningfully assessed in the continuous format, trials that had the same proportions as the counting consistent trials of the discrete formats were considered continuous "counting consistent" trials by convention. Similarly, continuous trials that showed the same proportions as the counting misleading trials were considered continuous "counting misleading" trials.

For all formats, we also manipulated the size of the individual spinners to prevent children from relying exclusively on the red area's absolute size in making their selections. Thus, on half of the trials of each format, the physically larger spinner also had the proportionally larger red area (size congruent trials). On the other half, the opposite pattern held, with the smaller spinner being the one with the proportionally larger red area (size incongruent trials). Spinners could be 6, 9, or 12 cm in diameter. For size congruent trials, the proportionally larger spinner was always the 12-cm spinner, and the other spinner could be 6 or 9 cm. In contrast, for size incongruent trials, the proportionally larger spinner was the 6-cm spinner. The other spinner was 9 or 12 cm. Proportion pairs (size congruent or size incongruent) were counterbalanced across participants. For all participants, the continuous condition was presented first, and the presentation order of the two discrete blocks was counterbalanced across participants. Importantly, children saw the same order in both time 1 and time 2 sessions.

For all conditions, trials started with a blank screen presented for 500 ms, followed by the pair of spinners. Spinners remained on the screen until the children responded by pressing one of two possible keys, "z" for the left spinner or "m" for the right one. Within each block, half of the correct responses were presented on the left and the other half on the right. More details about this task can be found at Abreu-Mendoza et al. (2020).

For consistency with the inhibitory control measure, we followed the same procedures for the H&F task when computing accuracy, which involved removing anticipatory and outlier responses. After applying these criteria, one participant who completed the intervention sessions did not have at least one trial from each type and was excluded from the final sample. Among the 49 children of the final sample, we analyzed 3429 (97.19%) of 3528 trials. For each participant, trial-level accuracy on this task was initially averaged by size (2), counting (2), format (3), and time (2), producing 24 data points per participant. However, given the complexity of the design and our theoretical interest in counting interference, representational format, and change over time, we then averaged across size, reducing the number of data points to 12 per participant. This approach provides a better estimate of performance within each trial type (e.g., counting misleading discrete adjacent at time 1) when there are unequal numbers of size congruent and incongruent trials (Abreu-Mendoza et al., 2020).

Intervention Program

The group-based proportional reasoning intervention program consisted of 24 one-hour sessions, which children attended twice a week. Throughout the sessions, students transitioned from representing proportions using manipulatives (i.e., Cuisenaire rods) to writing fraction expressions symbolically.

The intervention was divided into two phases: In Phase 1, children were introduced to the Cuisenaire rods, agreed on names for each different color rod, and a single letter to represent each rod color (usually the initial letter of the color name, see **Figure 1B**). This phase involved activities in which children internalized the correspondence between the rods' length and their colors and the relations of equality, inequality, and transitivity among the rods' lengths. For example, one activity involved asking children to close their eyes while a peer placed a rod in their hands and asked them to say aloud the color of the rod they are holding. Children were also asked to compare the different sizes of the rods relative to others (e.g., "the yellow rod is larger than the green rod") and place end-to-end, creating trains of rods of different sizes to equal the length of a larger rod. Then, children discussed the rods' lengths without having the rods present and verbally discussed relations among rod lengths and trains of rods. Later, using the letters to refer to rods, they wrote symbolic expressions such as "*y* is larger than *g*" or "*y* *g*." These tasks allowed children to move from non-symbolic to symbolic representations of lengths and relations among rod lengths. By the end of this phase, they could mentally evoke images of absolute magnitudes among rods of different lengths and symbolically represent those relations.

Phase 2 had four modules, which focused on the relative magnitude of the rods: In the first module, children were taught to use the rods as a tool to measure the length of other rods, which led to fractional expressions (e.g., "a white rod equals a third of a green rod"). Children described these relationships verbally with and then without the rods. In the second module, children used only symbols (i.e., operation and equality signs and letters) to refer to the rods and to write fractional expressions; for example, they might have written "$w = 1/3 \times g$." Children were then taught how to compare the proportional quantities of these expressions, which led to the final two modules. In the third module, children replaced letters with expressions that referred to the same magnitude; instead of writing *w*, they wrote expressions like the following "$1/3 \times g$." In the final module, children compared these symbolic fraction expressions (e.g., "$1/3 \times g < 5/7 \times e$").

During the two phases of the intervention, there were moment-by-moment formative assessments given by the instructors; however, there were no traditional summative assessments. Sessions were carried out by members of the research team: a university professor and a doctoral student. Further details of the program are reported in Powell (2019).

Statistical Analyses

Our first approach to examine training-related changes in the performance of children who participated in the intervention, was to contrast the performance of the control and the intervention groups across the three formats of the Spinner task, our outcome measure. **Table 1** shows the means and standard deviations for each group (control and intervention) across formats (continuous, discrete adjacent, and discrete mixed), counting conditions (consistent and misleading), and time (T1 and T2). For these three ANOVAs, we included time and counting as within-participant factors and group as

TABLE 1 | Mean (*SD*) performance across each format of the Spinners task by group.

	Intervention (*n* = 34)		Control (*n* = 15)	
	T1	**T2**	**T1**	**T2**
Continuous format				
Consistent	0.64 (0.25)	0.79 (0.18)	0.73 (0.23)	0.68 (0.21)
Misleading	0.64 (0.28)	0.69 (0.25)	0.71 (0.21)	0.74 (0.18)
Discrete adjacent format				
Consistent	0.79 (0.21)	0.91 (0.14)	0.78 (0.26)	0.80 (0.25)
Misleading	0.58 (0.25)	0.50 (0.25)	0.50 (0.30)	0.51 (0.29)
Discrete mixed format				
Consistent	0.85 (0.21)	0.86 (0.22)	0.77 (0.25)	0.88 (0.21)
Misleading	0.50 (0.29)	0.43 (0.26)	0.42 (0.34)	0.46 (0.33)

a between-participant factor. However, with one exception (a marginal three-way interaction between counting, time, and group ($p = 0.096$) in the continuous format), there were no time by group interaction to indicate greater learning in the original intervention group. Noting the irregular attendance of the intervention (1–24 sessions), we instead adopted a dose-response framework and performed retrospective analyses (Voils et al., 2014) on the effect of the number of sessions attended on changes in children's performance. Similar approaches have been employed to analyze results of educational interventions with incomplete attendance (Roberts et al., 2018), finding that when attrition rates are high and attendance is irregular, intervention effects might be better characterized by the mediator effects of the number of completed sessions.

Specifically, we computed linear mixed effects models using attendance as one of the between-participant fixed factors. To facilitate interpretation of results for all measures, we first standardized T1 and T2 scores relative to the T1 means and standard deviations. Therefore, all average performance at T1 measures would be centered around 0, and increments or decrements in the T2 session are reported in terms of standard deviations. Therefore, using the same guidelines as Cohen's d the resulting beta values can be interpreted using the follow criteria *small* 0.2 to < 0.5, *medium* 0.5 to < 0.8, and *large* 0.8 and above. Finally, to further characterize the main effects and interaction, we obtained the marginalized means at four levels of attendance (0, 8, 16, 24 sessions). We then performed pairwise comparisons at each level to determine if fitted increments or decrements differed significantly from zero. All statistical analyses were conducted in R 3.5.3 (R Core Team, 2019), linear mixed models were computed using the *lme4* package (Bates et al., 2015) and pairwise comparisons were performed using the *lsmeans* package (Lenth and Lenth, 2018). To investigate the effects of predictor variables, we first used Pearson's correlations between the learning gains in our outcome measure, the spinners task, across its different formats with the time 1 scores in the math, attitudinal, and inhibitory control measures. In cases where there was a significant correlation in the intervention group but not the control, we performed standard linear regressions with learning gains in the spinner task as the dependent variable, and examined interactions between number of sessions and the significant predictors to determine if the relations were specific to the training group. This approach also accounts for differences in sample size between the intervention and control group and hence the resulting power differences (Section "Participants and Intervention Phases").

RESULTS

Cognitive and Attitude Assessments

Pre-intervention Assessment

Table 2 shows the detailed descriptive statistics of the mathematical achievement, mathematical attitudes, and inhibitory control skills, before and after the intervention for both children who attended the intervention and children in the control group. Importantly, there were no differences in children's age, mathematical achievement, attitudes toward math, nor inhibitory control across the two groups before the intervention. Further, using attendance as a continuous variable there were no correlations of attendance with these demographic, academic, attitudinal, and cognitive variables (absolute $rs < 0.15$, $ps > 0.32$). Overall, these results suggest that any observed relationships between attendance and changes in proportional reasoning are not due to prior differences in math achievement or attitudes, or cognitive differences.

TABLE 2 | Group demographic, academic, and cognitive characteristics.

	Intervention group (n = 34)	Control (n = 15)	t(47)	p-value
Gender	15F/19M	6F/9M		
Pre-training				
Age	7.60 (0.38)	7.50 (0.27)	0.89	0.38
WIAT Math	81.12 (9.64)	80.80 (11.35)	0.10	0.92
ATM	4.05 (0.84)	4.29 (0.59)	1.00	0.32
H&F congruent	0.87 (0.12)	0.88 (0.14)	0.35	0.73
H&F incongruent	0.67 (0.31)	0.63 (0.28)	0.37	0.71
H&F mixed	0.55 (0.14)	0.50 (0.12)	1.33	0.19
Post-training				
Age	8.04 (0.37)	7.95 (0.26)	0.87	0.39
WIAT Math	88.18 (12.10)	88.00 (12.77)	0.05	0.96
ATM	4.24 (0.73)	4.15 (0.82)	0.38	0.70
H&F congruent	0.90 (0.12)	0.88 (0.13)	0.37	0.71
H&F incongruent	0.77 (0.26)	0.75 (0.24)	0.28	0.78
H&F mixed	0.59 (0.19)	0.54 (0.14)	1.05	0.30

H&F, Hearts and Flowers task; ATM, Attitudes Toward Math.

Time 1 to Time 2 Intervention Changes in Predictor Measures

To examine changes in the predictor measures over the five months between time 1 and time 2 and to determine whether attendance to the intervention program modulated these changes, we performed three linear mixed repeated models, which included time (T1 and T2) as a within-participant fixed factor and attendance as a between-participant fixed factor, and participants as a random factor with the standardized scores of each assessment as the dependent measure (**Figure 2**). For the H&F task, we also used block as a within-participant fixed categorical factor (congruent, incongruent, and mixed).

Mathematics achievement. There was a main effect of time (beta = 0.702, SE = 0.180, $t(47) = 3.91$, $p < 0.001$) for the arithmetic fluency scores of the WIAT-III. As we used the fall standard scores in time 1 and the spring standard scores for time 2, these results indicate that children improved their math skills above and beyond what was expected in a span of five months (**Figure 2A**). There was no main effect of attendance (beta = 0.015, SE = 0.017, $t(61.81) = 0.90$, $p = 0.370$) or interaction with attendance (beta < 0.001, SE = 0.012, $t(47) = 0.02$, $p = 0.984$), suggesting that these learning gains were independent of participating in the intervention program. All told, children's average scores shifted from low average (M = 81.02, SD = 10.07) to almost average (M = 88.12, SD = 12.17).

Positive Attitude toward Math. There was no main effect of time, nor attendance, nor interaction ($p > 0.320$), indicating that children's attitudes toward math remained stable from time 1 (M = 4.12, SD = 0.78) to time 2 (M = 4.22, SD = 0.75) regardless of their participation in the intervention program (**Figure 2B**).

Inhibitory control. Analyses of accuracy in the H&F task yielded a significant effect of incongruent block (beta = −0.864, SE = 0.215, $t(235)$ = 4.02, $p < 0.001$) and mixed block (beta = −1.424, SE = 0.215, $t(235)$ = 6.63, $p < 0.001$), indicating that children's performance was modulated by the block difficulty, with congruent, the easiest, followed by incongruent and then mixed (**Figure 2C**). There was also a marginal interaction between time and incongruent block (beta = 0.567, SE = 0.304, $t(235)$ = 1.87, p = 0.063), suggesting that there was a marginal moderate increase in children's performance in the incongruent block from time 1 to time 2. Notably, these improvements were not modulated by attendance to the intervention, with no main effect or interactions with attendance (all $ps > 0.35$).

Intervention Results

In the following sections, to provide evidence on the specificity of our results and rule-out test-re-test effects, we performed three linear mixed models, one for each format. We included the standardized accuracy scores (z-scores) as the dependent variable and added time (T1 vs. T2), counting (consistent and misleading) as within-participant fixed factors, and attendance as a between-participants factor, as well as the interactions, with participant as random intercept. Then, we calculated the marginal means, which were estimated from the linear mixed model, of the significant effects and interactions and performed pairwise comparisons of these estimated means to determine T1 and T2 differences across different number of completed sessions.

Continuous Proportional Reasoning

While there were no improvements in the control group, children in the intervention group improved in line with the number or sessions attended (**Figure 3**). The linear mixed model revealed a marginal interaction between time and attendance (beta = 0.033, SE = 0.198, $t(141)$ = 1.70, p = 0.091), confirming that children's improvements scaled with their attendance. To further characterize this interaction, we looked at the T1 and T2 marginalized mean differences at (0, 8, 16, and 24 sessions). These T1 vs. T2 comparisons indicated that while children who participated in zero sessions did not increase their performance (mean difference = 0.02, SE = 0.20, 95% CI [−0.415, 0.372], p = 0.913), children who received 8 (mean difference = 0.224, SE = 0.138, 95% CI [−0.046, 0.494], p = 0.107) and 16 sessions (mean difference = 0.470, SE = 0.151, 95% CI [0.174, 0.766], p = 0.0023) showed small increases. Finally, according to the model, children who completed the full intervention (24 sessions) had a large increment in their performance (mean difference = 0.716, SE = 0.227, 95% [0.271, 1.161], p = 0.002).

Outcomes in Discrete Proportional Reasoning

Discrete adjacent format. Children in the intervention group showed a decrease in their performance in the discrete adjacent misleading trials the more they attended to the intervention (**Figure 4**). This pattern of results was confirmed by the linear mixed model. This analysis yielded a negative linear effect of counting (beta = −0.963, SE = 0.251, $t(141)$ = 3.84, $p < 0.001$), which was qualified by a marginal three-way interaction between counting, time, and attendance (beta = −0.047, SE = 0.025, $t(141)$ = 1.92, p = 0.057). We performed two follow-up linear mixed models, one for each counting condition, with attendance and time as fixed factors and participants as random slopes. While the linear mixed model for the counting consistent condition did not yield any significant effect or interaction ($p > 0.21$), there was a marginal interaction between time and attendance in the misleading condition (beta = −0.036, SE = 0.018, $t(47)$ = 1.94, p = 0.058). Unexpectedly, this interaction indicated that children who attended more sessions decreased their performance in the misleading trials. The posthoc pairwise comparisons indicated that children who completed the full 24 sessions would be expected to have a significant decrease (beta = −0.652, SE = 0.298, 95% CI [−1.236, −0.068], p = 0.034), but not for those attending 16 or fewer sessions. In summary, these results indicate that the intervention might have increased the use of effective strategies for comparing discrete consistent trials but impair performance for cases in which there is a conflict between whole-number and proportional information.

Discrete mixed format. There were no meaningful changes in either of the two groups in the discrete mixed format (**Figure 5**). Consistently, the linear mixed model yielded a negative linear effect of counting (beta = −1.176, SE = 0.245, $t(141)$ = 4.80, $p < 0.001$), confirming the expected effect of lower performance on counting misleading problems. There was neither effect of time (p = 0.135) nor any two- nor three-way interactions involving time and attendance ($p > 0.20$).

Exploratory Analyses: The Relationship Between Academic, Attitudinal, and Cognitive Measures and Changes in Performance

To explore whether changes in proportional reasoning skills of children in the intervention group related to children's initial academic, cognitive and motivational skills, we performed Pearson correlations between the changes in the spinners task (T2 − T1) in continuous (**Figure 6**) and discrete adjacent (**Figure 7**) format and children's initial arithmetic fluency skills, attitudes toward math, and inhibitory control ability. We also examined the relationship between the developmental changes in proportional reasoning and academic, cognitive, and motivational skills in the control group. Importantly, we looked at these relationships separately as these predictor measures did not relate to each other (absolute r-values < 0.15, $ps > 0.32$). Finally, in cases where significant associations were found in the intervention group but not the control group, we formally tested for interactions using a linear regression analyses with the learning gains in proportional reasoning as the dependent variable, and the interaction between number of sessions and the significant predictor (math performance, attitudes toward math, or inhibitory control) as the moderator term.

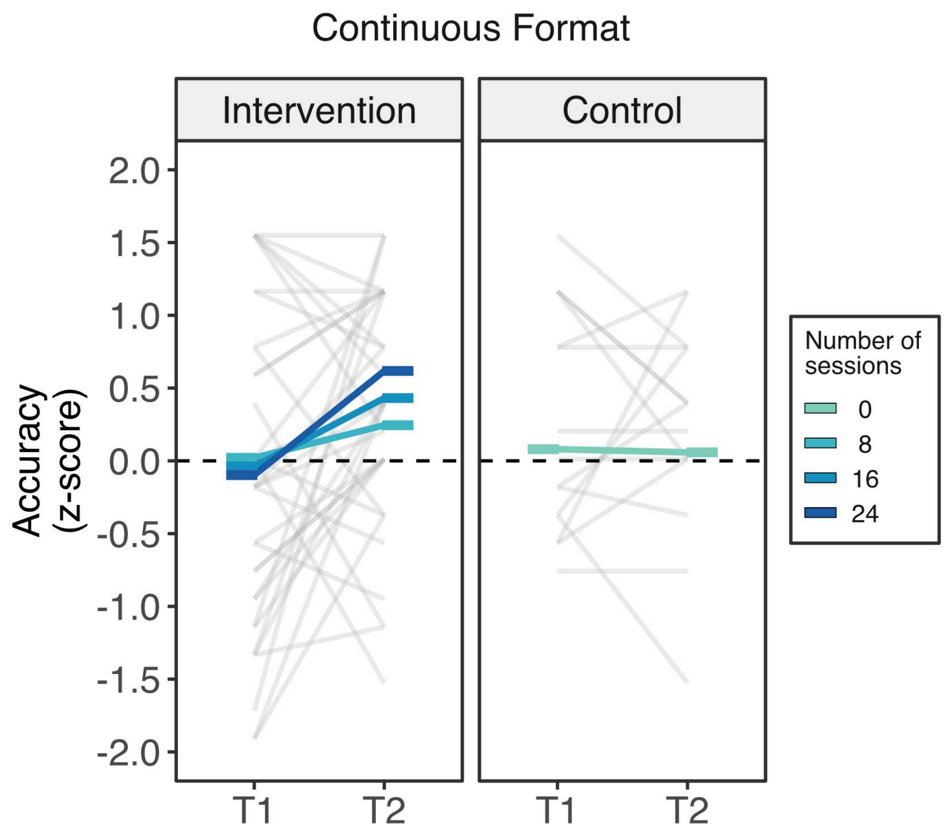


FIGURE 3 | Intervention outcomes in the non-symbolic continuous format of the Spinners task by the number of completed sessions. Children who attended all the sessions of the Cuisenaire intervention showed a reliable increase in their ability to compare spinners presented in the continuous format while children were part of the control group showed exhibited no improvement. Rods represents the estimated marginal means across different number of sessions of attendance from 0 sessions (control group) to 24 sessions (darkest blue).

Predictor Effects in Continuous Proportional Reasoning

After correcting for multiple comparisons using false discovery rate correction (Benjamini and Hochberg, 1995), there was a marginal association between the learning gains in the continuous format and T1 math scores ($r(32) = 0.36$, $p_{corr} = 0.057$) of children in the intervention group; that is, children with better math skills during the initial evaluation showed larger learning gains. We also found a significant positive relationship between attitudes toward math and children's learning gains ($r(32) = 0.41$, $p_{corr} = 0.046$), showing that children who had more positive attitudes toward math showed larger gains. Conversely, there was a marginal negative relationship between the time 1 score in the mixed block of the H&F task and the learning gains in non-symbolic continuous proportional reasoning skills ($r(32) = -0.29$, $p_{corr} = 0.095$). This pattern of results suggests that our intervention program resulted in larger gains for children with low-inhibitory control skills. In contrast, none of the correlations with change scores in the control group (all p_{corr}'s > 0.83).

A linear regression, with math performance as a moderator showed a significant interaction between math and attendance (beta = 0.0007, SE = 0.0003, $t = 2.08$, $p = 0.043$), suggesting that children who attended more sessions and had better math performance showed larger gains. There was also a marginal interaction between attendance and inhibitory control (beta = −0.046, SE = 0.027, $t = 1.69$, $p = 0.097$), suggesting that children with low inhibitory control but a higher number of completed sessions benefitted the most from the intervention. Finally, there was no interaction between attendance and attitude toward math scores ($p = 0.29$).

Predictor Effects in Discrete Proportional Reasoning

There was a negative correlation for discrete proportional reasoning for misleading trials. Children in the intervention group with lower inhibitory control had larger gains in misleading trials ($r(32) = -0.41$, $p_{corr} = 0.045$). There were no significant correlations between individual difference measures and changes in the discrete adjacent misleading trials in the control group (all p_{corr}'s > 0.27).

The linear regression to predict learning gains in the discrete adjacent misleading condition with attendance and inhibitory control as independent variables did not yield a significant interaction between these two factors ($p = 0.34$).

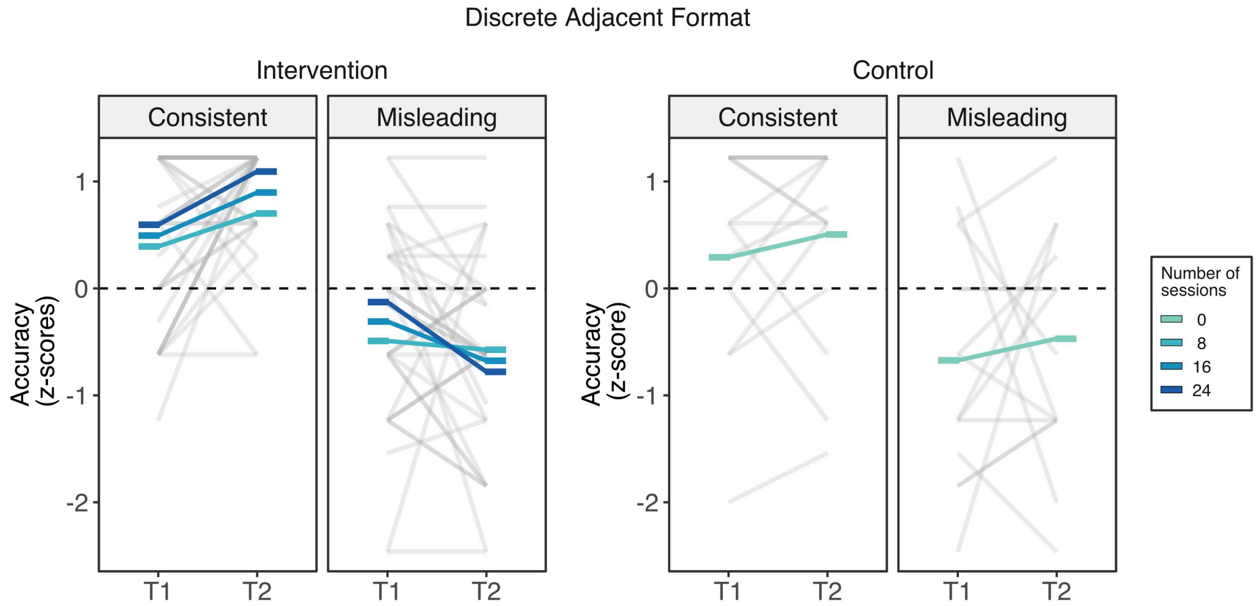


FIGURE 4 | Intervention outcomes in the non-symbolic discrete adjacent format of the Spinners task across groups. Unexpectedly, although improvements in the ability to compare pairs of spinners whose number of segments matched the proportional information varied across did not vary by the number of completed session, children's ability to compare spinners when the number of segments was misleading was negatively related to the number of sessions. That is, children who completed 24 session (darkest blue line) showed the largest losses. Rods represents the estimated marginal means across different number of sessions of attendance from 0 sessions (control group) to 24 sessions (darkest blue).

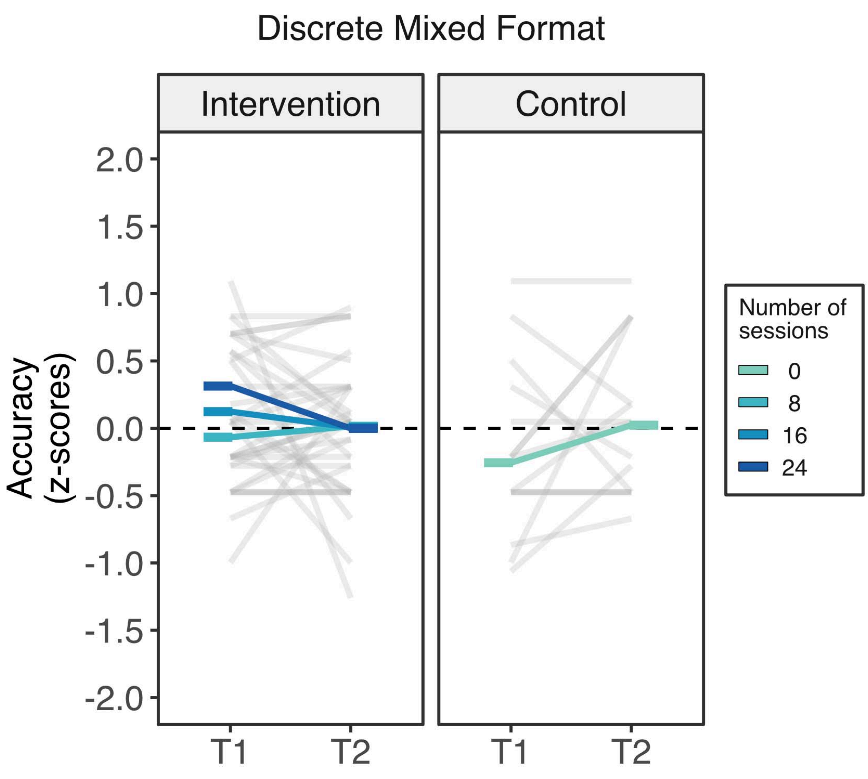


FIGURE 5 | Intervention outcomes in the non-symbolic discrete mixed format of the Spinners task across groups. Changes in the discrete mixed format from time 1 to time 2 did not vary by attendance to the intervention. Rods represents the estimated marginal means across different number of sessions of attendance from 0 sessions (control group) to 24 sessions (darkest blue).

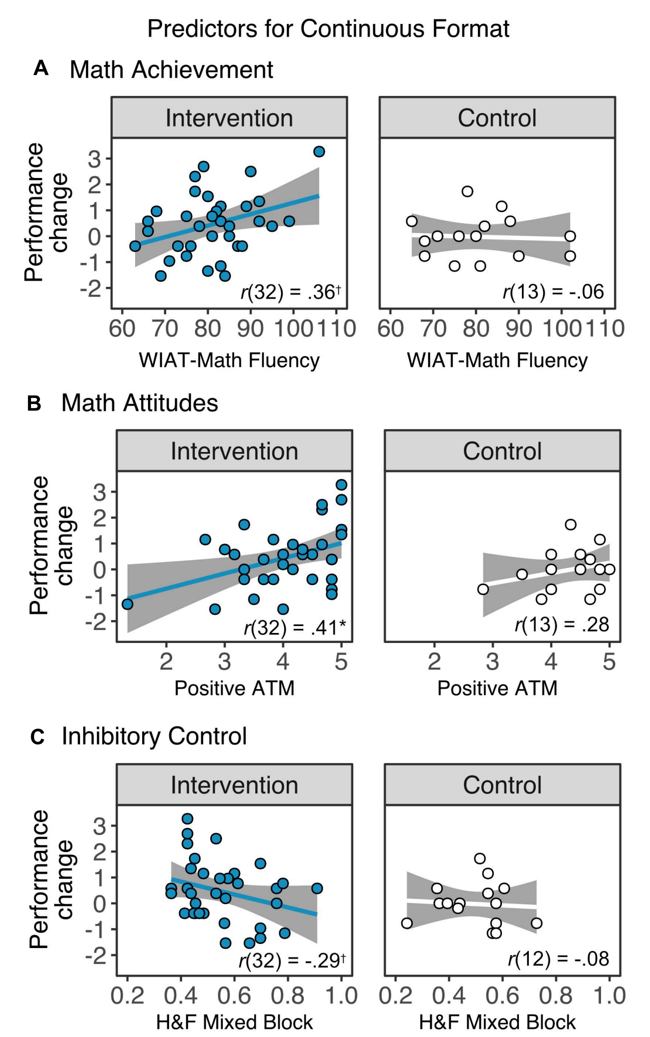


FIGURE 6 | Associations between changes in performance in the continuous format of the Spinners task and **(A)** math achievement, as measured by the fluency subtests from the Wechsler Individual Achievement Test (WIAT)-III, **(B)** attitudes toward math (ATM), and **(C)** inhibitory control, as measured by the mixed block of the Hearts and Flowers (H&F) task. Children in the intervention group with better math skills and more positive attitudes toward math showed marginally larger learning gains. In contrast, children with lower inhibitory control skills in this group showed marginal larger gains. Notably, none of the correlations with change scores in the control group reached significance. $^{*}p_{corr} < 0.05$, $^{\dagger}p_{corr} < 0.10$.

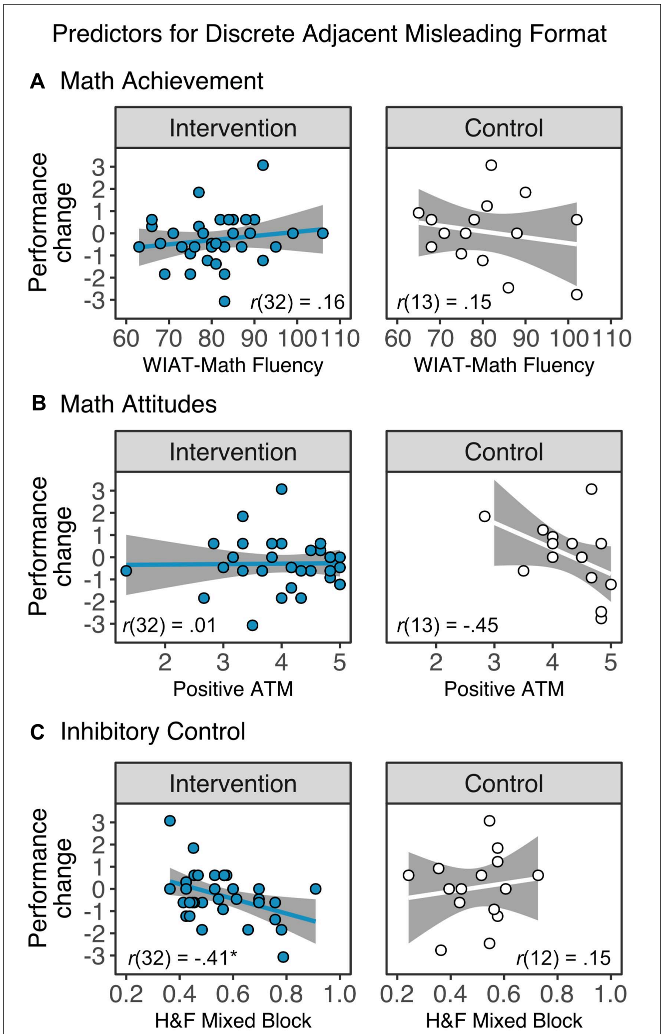


FIGURE 7 | Associations between changes in performance in discrete adjacent misleading trials of the Spinners task and **(A)** math achievement, as measured by the fluency subtests from the Wechsler Individual Achievement Test (WIAT)-III, **(B)** attitudes toward math (ATM), and **(C)** inhibitory control, as measured by the Hearts and Flowers (H&F) task. There were no significant correlations between individual difference measures and changes in the discrete adjacent misleading trials, except for a negative association between the initial inhibitory control of the intervention group and learning gains. $^{*}p_{corr} < 0.05$.

DISCUSSION

Training studies have the potential of uncovering the mechanisms that underpin effective learning programs (Rosenberg-Lee, 2018). Emerging research has pointed to non-symbolic proportional reasoning as a building block for later fraction understanding (Siegler et al., 2011; Matthews et al., 2016). In the current study, we examined the malleability of this skill by looking at the effects of a 24-session proportional reasoning intervention (Powell, 2019), using Cuisenaire rods, on second-grade students' ability to compare non-symbolic continuous and discrete ratios. We found that children who went through essentially the complete intervention showed a large increase in their ability to compare non-symbolic continuous proportions in a different representational format, annulus-shaped figures. However, we observed a decline in their ability to compare discretized proportions, specifically, when the absolute number of pieces contradicted the proportional information. In contrast, while children who did not attend any sessions did not show any improvement or decline in their proportional reasoning skills, even after the five months, children who received at least 16 sessions showed a small but consistent increment in their continuous proportional skills. Finally, we found a positive predictor role of children's aptitudes and attitudes in mathematics on the continuous learning outcomes, but a negative role for children's inhibitory control skills.

Cuisenaire Rod Intervention Improves Non-symbolic Continuous Proportional Reasoning

Children who completed the 24 sessions of our intervention showed an increase in their non-symbolic continuous proportional reasoning. These results are consistent with previous findings showing that non-symbolic proportional reasoning is malleable through training non-symbolic estimation skills (Gouet et al., 2020) and mapping between non-symbolic and symbolic formats (Wortha et al., 2020). Importantly, in contrast to previous interventions implemented in highly structured, individual, computerized environments, the current intervention took place in an ecologically valid context (small group instruction in the children's classroom) and used inexpensive materials (Cuisenaire rods). Further, our approach was implemented slowly over five months compared to only a few days in prior work (e.g., Gouet et al., 2020). These features contribute to the ecological validity of our intervention, which could facilitate its implementation by teachers. Conversely, these features may complicate implementing this approach in a large scale, randomized control trial needed to establish efficacy.

Our study also expands these previous findings by showing that children can transfer their gains in proportional reasoning to a different non-symbolic representation after an intervention focused explicitly on conceptual thinking. In the current study, children extended their recently acquired understanding of proportional reasoning with one model (lines) to one in which they did not receive any practice (annulus-shaped figures). Significantly, these results cannot be attributed to test-retest effects as children who did not attend any session did not show any improvement.

During the intervention, children viewed proportions represented in both non-symbolic and symbolic formats. Working with these two formats might have facilitated transferring proportional reasoning skills from one modality (rods) to another (annulus-shaped figures). Children compared the relative lengths of rods and used verbal and written symbolic expressions to represent them. Thus, children might have developed representations of proportions that were not constrained by the perceptual properties of the stimuli they were manipulating. Consistent with this interpretation, previous fraction interventions have shown that using different non-symbolic representations of proportions (e.g., circles and tiles) helps children transfer gains in fraction comparison ability to fraction magnitude estimation skills in children with low working memory (Fuchs et al., 2014). Future studies should examine whether the non-symbolic or the symbolic features of the intervention or the combination of both, drove children's improvements in proportional reasoning.

Discrete Proportional Reasoning Is Hindered by the Intervention

In contrast to the relative ease of processing continuous proportions, discrete quantities pose a significant challenge to non-symbolic proportional reasoning. Students not only have to manipulate the proportional quantities but also have to override the whole-number information (Jeong et al., 2007; Boyer et al., 2008; Abreu-Mendoza et al., 2020). Two lines of research suggest that changes in continuous proportional reasoning should also lead to improvements in discrete reasoning. First, there is an emergent body of research suggesting that priming non-symbolic continuous proportions leads to short-term improvements in processing discrete quantities (Boyer and Levine, 2015; Hurst and Cordes, 2018; Abreu-Mendoza et al., 2020). Second, a recent proposal has suggested a cognitive system devoted to processing proportional quantities regardless of format and modality (Matthews et al., 2016). Therefore, we hypothesized that by training proportional reasoning with a particular focus on a continuous magnitude (the relative lengths of Cuisenaire rods), children would not only improve their ability to compare continuous proportions but also discrete ones. However, contrary to this hypothesis, children who completed the 24 sessions of the intervention showed a decline in their ability to compare discrete quantities, specifically in contexts where whole-number information interfered with the proportional one. These results beg the question if proportions are processed in a modality-independent manner (Matthews et al., 2016; Park et al., 2020), why do gains in continuous proportional reasoning not transfer to discrete quantities? We offer two potential explanations: one related to our intervention's instructional structure and the other to the developmental trajectory of discrete proportional reasoning.

One feature of our intervention that may have hindered children's discrete reasoning skills is the nature of the Phase 1 activities. Some of these activities involved combining two or more rods and contrasting them with rods of larger lengths. Although these activities focused on a continuous magnitude, length, children still might have focused on discrete elements (rods). Moreover, comparing proportions in the context of whole-number interference was not an explicit topic of the intervention. This omission may have been compounded by features of the intervention aimed at increasing student agency: participants proposed the proportion problems they would work on during the sessions; thus, students may have never encountered counting misleading problems (Rosenberg-Lee, 2021). These two features of the intervention combined with the on-going instruction of whole-number operations and units that children received during the second grade (Common Core State Standards Initiative, 2020a) might have led children to focus on the absolute number of pieces instead of the proportional information in our discrete outcome task.

The protracted development of discrete proportional reasoning might have also played a part in children's decrease in this ability. For instance, a recent study found that even fifth-graders with low-fraction knowledge performed at chance level in a match-to-sample task that involved discrete-adjacent stimuli (Begolli et al., 2020). Only children with at least a moderate level of fraction knowledge could manipulate discrete quantities, suggesting that *symbolic* proportional skills might be required to overcome the misleading discrete information while manipulating non-symbolic proportional quantities (Begolli et al., 2020). The authors suggest that children might need more symbolic proportional experience to disregard discrete information successfully.

In summary, consistent with these previous studies, our current findings suggest that to help children in the protracted development of discrete proportional reasoning (Jeong et al., 2007), children may require direct instruction that brings to conscious attention the interference of whole-number knowledge to proportional reasoning (Rosenberg-Lee, 2021), explicitly linking continuous and discrete proportional reasoning and a more extensive period of symbolic proportional instruction.

Math Achievement and Inhibitory Control Moderate Learning Outcomes

The third goal of this study was to examine the predictive relations of measures of math aptitude, math attitude, and inhibitory control abilities with changes in performance in non-symbolic continuous and discrete proportional reasoning skills. Despite our small sample size, but consistent with longitudinal studies of fractions (Jordan et al., 2013) and general math growth (Dowker et al., 2019), both initial math performance and attitudes toward math showed a positive relationship with children's learning outcomes in the intervention group. Notably, only math performance had a positive interaction with attendance when predicting learning gains in continuous proportional reasoning, suggesting that children with higher math skills and that attended more of the intervention showed the greatest gains in their ability to compare proportions of continuous quantities.

Conversely, children's initial inhibitory control skills in the intervention group were negatively related to their learning gains in continuous proportional reasoning and losses in discrete proportional reasoning. Further, inhibitory control had a negative interaction with attendance in predicting gains in the continuous format. That is, children with low inhibitory control but with high attendance benefit the most from the intervention. This pattern of results might appear counterintuitive at first sight, especially in light of previous studies reporting a positive relationship between inhibitory control and non-symbolic (Abreu-Mendoza et al., 2020) and symbolic proportional processing (Gomez et al., 2015; Avgerinou and Tolmie, 2019; Coulanges et al., 2021). However, a parsimonious interpretation of these results is that the current intervention is beneficial to remediate the proportional reasoning skills of those students with low inhibitory control skills. Indeed, a previous study examining aptitude-treatment interactions in the context of fraction learning reported a similar result: a conceptually rich program was more effective to improve fraction knowledge of children with low working memory, than a fraction fluency intervention (Fuchs et al., 2014). However, the current study leaves unanswered whether this effect is specific to the assessed executive function (i.e., inhibitory control). Future studies contrasting different non-symbolic proportional reasoning programs that include a more comprehensive cognitive assessment would provide insights on these questions. Further, it raises the possibility that exposing children with poor inhibitory control to proportional stimuli may help them to build up non-symbolic representation that could be helpful for fraction processing (Matthews et al., 2016; Kalra et al., 2020). Relating changes in proportional reasoning to improvements in symbolic fraction understanding, especially among learners with inhibitory control deficits, is vital in understanding the mechanisms supporting this remediation approach (Rosenberg-Lee, 2018).

Considerations and Future Directions

While recent interventions on non-symbolic and symbolic fraction knowledge have involved individual repetition of computerized tasks (Fazio et al., 2016; Gouet et al., 2020), the current intervention program comprised tasks that involved children working in small groups, using well-known educational materials, Cuisenaire rods. This ecologically valid implementation allowed children to build a conceptual understanding of proportions by first working with whole non-symbolic and symbolic magnitudes (first phase) and slowly transition to non-symbolic and symbolic proportions (second phase). However, the intervention's cumulative nature, which required students to attend all sessions, make children's irregular and unsystematic attendance a significant limitation of the study. For example, while some children completed the eight sessions of the first phase but did not attend any of the second phase session, others attended 15 out of the 16 sessions of the second phase but none of the first one, making it difficult to disentangle the effects of the number of sessions from to content of the session. Our results indicated that children who attended the complete intervention showed the largest changes in their performance; however, only six out of the 34 children of the intervention group attended the complete intervention (24 sessions), and only 14 children attended 80% or more sessions. This low rate in completing the intervention might have contributed to the marginal significance of some of the results. Larger samples are needed to verify pattern of results especially gains in continuous but losses in discrete misleading. While short-term computerized instruction can produce gains (Fazio et al., 2016; Gouet et al., 2020), conceptual instruction has been shown to have long-standing positive impacts (Misquitta, 2011). Future research should contrast the cost-benefits of long-term, group-based, conceptual instruction and short-term, individual interventions on children's proportional reasoning.

CONCLUSION

The current study examined the effects of a proportional reasoning intervention, in which children transition from non-symbolic to symbolic fraction comparisons and expressions, on children's ability to compare ratios of continuous and discrete quantities. Our results showed that children who completed the intervention increased their ability to compare non-symbolic continuous proportions. However, contrary to our expectations, children decreased their ability to compare misleading discretized proportions. These results provide further evidence on the malleability of non-symbolic continuous proportional reasoning and speak to the persistence of difficulties with discrete proportional skills.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because participants did not consent to data sharing. Requests to access the datasets should be directed to MR-L, miriam.rosenberglee@rutgers.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Rutgers University Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

RA-M, KA, AP, and MR-L conceived the study. RA-M and LC collected the data. RA-M performed the analyses. RA-M, AP, and MR-L wrote the original draft. RA-M, LC, AP, and MR-L reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

AP was supported by the National Council of Teacher of Mathematics under the Mathematics Education Trust (MET) 2018–2019 Classroom Research Grant. Elisee Doh gratefully acknowledges receiving a research stipend from the National Science Foundation through the Garden State-LSAMP (NSF Award 1909824).

ACKNOWLEDGMENTS

We thank the principal, Simone Rose, as well as Shirlene Crockett and all the teachers and staff at Thirteenth Avenue/MLK, Jr. School who helped in this study. We also thank Blessing Eluwa, Milagros Grados, Jessica Amato, Dima Karim, Isabel Paucar, Kainaat Anwar, and Catalina Guzman for their assistance with data collection as well as Elisee Doh for stimuli generation.

REFERENCES

Abreu-Mendoza, R. A., Coulanges, L., Ali, K., Powell, A. B., and Rosenberg-Lee, M. (2020). Children's discrete proportional reasoning is related to inhibitory control and enhanced by priming continuous representations. *J. Exp. Child Psychol.* 199:104931. doi: 10.1016/j.jecp.2020.104931

Avgerinou, V. A., and Tolmie, A. (2019). Inhibition and cognitive load in fractions and decimals. *Br. J. Educ. Psychol.* 90(Suppl. 1), 240–256. doi: 10.1111/bjep.12321

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.

Begolli, K. N., Booth, J. L., Holmes, C. A., and Newcombe, N. S. (2020). How many apples make a quarter? The challenge of discrete proportional formats. *J. Exp. Child Psychol.* 192:104774. doi: 10.1016/j.jecp.2019.104774

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bhatia, P., Delem, M., Léone, J., Boisin, E., Cheylus, A., Gardes, M.-L., et al. (2020). The ratio processing system and its role in fraction understanding: evidence from a match-to-sample task in children and adults with and without dyscalculia. *Q. J. Exp. Psychol.* 73, 2158–2176. doi: 10.1177/1747021820940631

Boyer, T. W., and Levine, S. C. (2015). Prompting children to reason proportionally: processing discrete units as continuous amounts. *Dev. Psychol.* 51, 615–620. doi: 10.1037/a0039010

Boyer, T. W., Levine, S. C., and Huttenlocher, J. (2008). Development of proportional reasoning: where young children go wrong. *Dev. Psychol.* 44, 1478–1490. doi: 10.1037/a0013110

Braithwaite, D. W., and Siegler, R. S. (2020). Putting fractions together. *J. Educ. Psychol.* 113, 556–571. doi: 10.1037/edu0000477

Brocki, K. C., and Tillman, C. (2014). Mental set shifting in childhood?: The role of working memory and inhibitory control. *Infant and Child Development* 604(June), 588–604. doi: 10.1002/icd

Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., and Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS One* 13:e0198973. doi: 10.1371/journal.pone.0198973

Carpenter, T., Corbitt, M., Kepner, H., Lindquist, M., and Reys, R. (1980). Results of the second NAEP mathematics assessment: secondary school. *Math. Teacher* 73, 329–338. doi: 10.5951/mt.73.5.0329

Chen, L., Bae, S. R., Battista, C., Qin, S., Chen, T., Evans, T. M., et al. (2018). Positive attitude toward math supports early academic success: behavioral evidence and neurocognitive mechanisms. *Psychol. Sci.* 29, 390–402. doi: 10.1177/0956797617735528

Common Core State Standards Initiative (2020a). *Measurement & Data*. Corestandards. Available online at: http://www.corestandards.org/Math/Content/2/NBT/ (accessed November 16, 2020).

Common Core State Standards Initiative (2020b). *Number & Operations—Fractions*. Corestandards. Available online at: http://www.corestandards.org/Math/Content/NF/ (accessed October 07, 2020).

Coulanges, L., Abreu-Mendoza, R. A., Varma, S., Uncapher, M., Gazzaley, A., Anguera, J., et al. (2021). Linking inhibitory control to math achievement via comparison of conflicting decimal numbers. *Cognition*.

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.3928/0090-4481-20050401-12

Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44, 2037–2078.

Dowker, A., Cheriton, O., Horton, R., and Mark, W. (2019). Relationships between attitudes and performance in young children's mathematics. *Educ. Stud. Math.* 100, 211–230. doi: 10.1007/s10649-019-9880-5

Fazio, L. K., Kennedy, C. A., and Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PLoS One* 11:e0165243. doi: 10.1371/journal.pone.0165243

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., et al. (2013). Improving at-risk learners' understanding of fractions. *J. Educ. Psychol.* 105, 683–700. doi: 10.1037/a0032446

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., et al. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *J. Educ. Psychol.* 106, 499–514. doi: 10.1037/a0034341

Gomez, D. M., Jimenez, A., Bobadilla, R., Reyes, C., and Dartnell, P. (2015). The effect of inhibitory control on general mathematics achievement and fraction comparison in middle school children. *ZDM Int. J. Math. Educ.* 47, 801–811. doi: 10.1007/s11858-015-0685-4

Gouet, C., Carvajal, S., Halberda, J., and Peña, M. (2020). Training nonsymbolic proportional reasoning in children and its effects on their symbolic math abilities. *Cognition* 197:104154. doi: 10.1016/j.cognition.2019.104154

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

Hurst, M. A., and Cordes, S. (2018). Attending to relations: proportional reasoning in 3- to. *Dev. Psychol.* 54, 428–439. doi: 10.1037/dev0000440

Jeong, Y., Levine, S. C., and Huttenlocher, J. (2007). The development of proportional reasoning: effect of continuous versus discrete quantities. *J. Cogn. Dev.* 8, 237–256. doi: 10.1080/15248370701202471

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

Kalra, P. B., Binzak, J. V., Matthews, P. G., and Hubbard, E. M. (2020). Symbolic fractions elicit an analog magnitude representation in school-age children. *J. Exp. Child Psychol.* 195:104844. doi: 10.1016/j.jecp.2020.104844

Lenth, R., and Lenth, M. R. (2018). Package 'lsmeans.'. *Am. Stat.* 34, 216–221.

Lortie-Forgues, H., Tian, J., and Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Dev. Rev.* 38, 201–221. doi: 10.1016/j.dr.2015.07.008

Mathematical Cognition and Learning Society (2019). *2019 Mathematical Cognition and Learning Society Conference*. Available online at: https://www.the-mcls.org/fileadmin/user_upload/MCLS_Full_program_.pdf (accessd November 16, 2020).

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

Misquitta, R. (2011). A review of the literature: fraction instruction for struggling learners in mathematics. *Learn. Disabil. Res. Pract.* 26, 109–119. doi: 10.1111/j.1540-5826.2011.00330.x

Mock, J., Huber, S., Bloechle, J., Dietrich, J. F., Bahnmueller, J., Rennig, J., et al. (2018). Magnitude processing of symbolic and non - symbolic proportions: an fMRI study. *Behav. Brain Funct.* 14:9. doi: 10.1186/s12993-018-0141-z

Ni, Y., and Zhou, Y. (2005). Teaching and learning fraction and rational numbers: the origins and implications of whole number bias. *Educ. Psychol.* 40, 27–52. doi: 10.1207/s15326985ep4001_3

Park, Y., Viegut, A. A., and Matthews, P. G. (2020). More than the sum of its parts: exploring the development of ratio magnitude vs. simple magnitude perception. *Dev. Sci.* e13043. doi: 10.1111/desc.13043

Pierce, J. W. (2007). PsychoPy- psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Plummer, P., DeWolf, M., Bassok, M., Gordon, P. C., and Holyoak, K. J. (2017). Reasoning strategies with rational numbers revealed by eye tracking. *Attention Percept. Psychophys.* 79, 1426–1437. doi: 10.3758/s13414-017-1312-y

Powell, A. B. (2019). Aprimorando O conhecimento dos estudantes sobre a magnitude da fração: um estudo preliminar com alunos nos anos inicias. *Rev. Int. Pesq. Educ. Matemátic.* Xiii, 50–68.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing, Reference Index Version 3.5.3*.

Reinhold, F., Hoch, S., Werner, B., Richter-Gebert, J., and Reiss, K. (2020). Learning fractions with and without educational technology: what matters for high-achieving and low-achieving students? *Learn. Instruction* 65:101264. doi: 10.1016/j.learninstruc.2019.101264

Roberts, G. J., Capin, P., Roberts, G., Miciak, J., Quinn, J. M., and Vaughn, S. (2018). Examining the effects of afterschool reading interventions for upper elementary struggling readers. *Remedial Special Educ.* 39, 131–143. doi: 10.1177/0741932517750818

Rosenberg-Lee, M. (2021). "Uncovering the neural basis rational number difficulties: the role of inhibitory control and magnitude processing," in *Heterogeneous Contributions to Numerical Cognition. Learning and Education in Numerical Cognition*, eds W. Fias and A. Henik (Academic Press).

Rosenberg-Lee, M. (2018). Training studies: an experimental design to advance educational neuroscience. *Mind Brain Educ.* 12, 12–22. doi: 10.1111/mbe.12166

Sidney, P. G., Thompson, C. A., Fitzsimmons, C., and Taber, J. M. (2019). Children's and adults' math attitudes are differentiated by number type. *J. Exp. Educ.* 89, 1–32. doi: 10.1080/00220973.2019.1653815

Siegler, R. S., and Lortie-Forgues, H. (2017). Hard lessons: why rational number arithmetic is so difficult for so many people. *Curr. Direct. Psychol. Sci.* 26, 346–351. doi: 10.1177/0963721417700129

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

Soni, M., and Okamoto, Y. (2020). Improving children's fraction understanding through the use of number lines. *Math. Think. Learn.* 22, 233–243. doi: 10.1080/10986065.2020.1709254

van Hoof, J., Degrande, T., Ceulemans, E., Verschaffel, L., and van Dooren, W. (2018). Towards a mathematically more correct understanding of rational numbers: a longitudinal study with upper elementary school learners. *Learn. Individ. Differ.* 61, 99–108. doi: 10.1016/j.lindif.2017.11.010

Voils, C. I., King, H. A., Maciejewski, M. L., Allen, K. D., Yancy, W. S., and Shaffer, J. A. (2014). Approaches for informing optimal dose of behavioral interventions. *Ann. Behav. Med.* 48, 392–401. doi: 10.1007/s12160-014-9618-7

Vosniadou, S., Pnevmatikos, D., Makris, N., Lepenioti, D., Eikospentaki, K., Chountala, A., et al. (2018). The Recruitment of shifting and inhibition in on-line science and mathematics tasks. *Cogn. Sci.* 42, 1860–1886. doi: 10.1111/cogs.12624

Wechsler, D. (2009). *Wechsler Individual Achievement Test–Third Edition*. London: Pearson.

Wortha, S. M., Bloechle, J., Ninaus, M., Kiili, K., Lindstedt, A., Bahnmueller, J., et al. (2020). Neurofunctional plasticity in fraction learning: an fMRI training study. *Trends Neurosci. Educ.* 21:100141. doi: 10.1016/j.tine.2020.100141

Wright, A., and Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Front. Psychol.* 5(March), 1–9. doi: 10.3389/fpsyg.2014.00213

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 29 June 2021
doi: 10.3389/fpsyg.2021.649109

Check for updates

Dysnumeria in Sign Language: Impaired Construction of the Decimal Structure in Reading Multidigit Numbers in a Deaf ISL Signer

Naama Friedmann*, Neta Haluts and Doron Levy

Language and Brain Lab, Tel Aviv University, Tel Aviv, Israel

We report on the first in-depth analysis of a specific type of dysnumeria, number-reading deficit, in sign language. The participant, Nomi, is a 45-year-old signer of Israeli Sign Language (ISL). In reading multidigit numbers (reading-then-signing written numbers, the counterpart of reading aloud in spoken language), Nomi made mainly decimal, number-structure errors– reading the correct digits in an incorrect (smaller) decimal class, mainly in longer numbers of 5–6-digits. A unique property of ISL allowed us to rule out the numeric-visual analysis as the source of Nomi's dysnumeria: In ISL, when the multidigit number signifies the number of objects, it is signed with a decimal structure, which is marked morphologically (e.g., 84 → Eight-Tens Four); but a parallel system exists (e.g., for height, age, bus numbers), in which multidigit numbers are signed non-decimally, as a sequence of number-signs (e.g., 84 → Eight, Four). When Nomi read and signed the exact same numbers, but this time non-decimally, she performed significantly better. Additional tests supported the conclusion that her early numeric-visual abilities are intact: she showed flawless detection of differences in length, digit-order, or identity in same-different tasks. Her decimal errors did not result from a number-structure deficit in the phonological-sign output either (no decimal errors in repeating the same numbers, nor in signing multidigit numbers written as Hebrew words). Nomi had similar errors of conversion to the decimal structure in number comprehension (number-size comparison tasks), suggesting that her deficit is in a component shared by reading and comprehension. We also compared Nomi's number reading to her reading and signing of 406 Hebrew words. Nomi's word reading was in the high range of the normal performance of hearing controls and of deaf signers and significantly better than her multidigit number reading, demonstrating a dissociation between number reading, which was impaired, and word reading, which was spared. These results point to a specific type of dysnumeria in the number-frame generation for written multidigit numbers, whereby the conversion from written multidigit numbers to the abstract decimal structure is impaired, affecting both reading and comprehension. The results support abstract, non-verbal decimal structure generation that is shared by reading and comprehension, and also suggest the existence of a non-decimal number-reading route.

Keywords: number impairment, sign language, number reading, dyscalculia, deaf, number reading model, reading, dysnumeria

Edited by:
Júlia Beatriz Lopes-Silva, Federal University of Minas Gerais, Brazil

Reviewed by:
Chloe Marshall, University College London, United Kingdom
Bencie Woll, University College London, United Kingdom

***Correspondence:**
Naama Friedmann
naamafr@tauex.tau.ac.il

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

Received: *06 January 2021*
Accepted: *29 April 2021*
Published: *29 June 2021*

Citation:
Friedmann N, Haluts N and Levy D (2021) Dysnumeria in Sign Language: Impaired Construction of the Decimal Structure in Reading Multidigit Numbers in a Deaf ISL Signer. Front. Psychol. 12:649109. doi: 10.3389/fpsyg.2021.649109

1. INTRODUCTION

Number reading, just like word reading, is a complex, multi-staged process (McCloskey et al., 1985, 1986, 1990; Cohen et al., 1997; Dehaene et al., 2003; Dotan and Friedmann, 2018), which is essential in our everyday life (Nuerk et al., 2015). A deficit in each of the components of this number-reading process gives rise to a different type of dysnumeria, which manifests itself in different types of errors and in different patterns of performance in various number tasks (McCloskey et al., 1985, 1986, 1990; Temple, 1989; Noel and Seron, 1993; Cipolotti and Butterworth, 1995; Cipolotti et al., 1995; Cohen et al., 1997; Basso and Beschin, 2000; Deloche and Willmes, 2000; Delazer and Bartha, 2001; Dehaene et al., 2003; Cappelletti et al., 2005; Friedmann et al., 2010; Starrfelt et al., 2010; Starrfelt and Behrmann, 2011; Moura et al., 2013; Dotan and Friedmann, 2015, 2018). Until now, dysnumerias have been reported for spoken languages. In this paper we report on the first in-depth analysis of a specific dysnumeria in sign language, in Nomi, a 45-year-old signer of Israeli Sign Language (ISL).

Previous studies reported that compared to hearing individuals, deaf individuals have difficulties with numbers, mostly in arithmetic and mathematics (Wollman, 1964; Austin, 1975; Wood et al., 1984; Titus, 1995; Frostad, 1996; Nunes and Moreno, 1998; Traxler, 2000; Davis and Kelly, 2003; Bull et al., 2011; Gottardis et al., 2011). However, these studies referred to the deaf population as a whole, without examining signers specifically, and they referred to perfomance in general mathematics tests, without assessing number-reading. Additionally, these reports are mainly obtained from general math tests administered in English, and these have been associated with difficulties in the spoken language (cf. Kelly and Gaustad, 2007, where the performance in mathematics tests was found to correlate with the performance in reading and morphology in English). Other studies have focused on specific properties of sign languages and their effects on number processing, e.g., how the sub-base 5 of the numeral system in the German sign language, Deutsche Gebärdensprache (DGS) affects parity retrieval (Iversen et al., 2006), and how DGS properties affect the response times on parity judgmenets with the left or right hand (MARC effect, Iversen et al., 2004), but have not related to specific difficulties in number processing. The studies that did assess number-reading difficulties in deaf individuals have not tested signing of multidigit numbers (Genovese et al., 2005; Korvorst et al., 2007; Domahs et al., 2010, 2012; Palma et al., 2010). As a consequence, no study has analyzed the pattern of errors made by deaf signers in reading multidigit numbers, and there have been no reports of selective impairments in number-reading in sign language. However, as we show below, testing dysnumeria in sign language offers interesting insights to dysnumeria and to the number-reading model.

1.1. The Number Reading Process

Dotan and Friedmann (2018) proposed an intergrated model for reading aloud of numbers (depicted in **Figure 1**), which combines elements from the triple-code model (Dehaene, 1992; Dehaene and Cohen, 1995; Dehaene et al., 2003), and the number-reading models by McCloskey and colleagues (McCloskey et al., 1986; McCloskey, 1992) and Cohen and Dehaene (1991), and refines them based on findings from neuropsychological case studies of specific number impairments.

According to the model (**Figure 1**), multidigit number reading begins with numeric-visual analysis of the written number, which includes separate mechanisms for encoding of digit-identity, digit-order, and the extraction of decimal structure. The process of decimal structure extraction includes separate mechanisms for encoding of the number-length, detection of zeros and their positions within the number, and parsing of triplets.

The information about digit-identity and digit-order is then held in a dedicated numeric buffer as ordered digits, and the information on the decimal structure is sent to components responsible for creation of the number-word frame—the verbal form of the number. First, a syntactic tree is built, which is a hierarchical representation of the number (for example, 3-digit numbers are represented by a node for tens, a node for ones, and a higher node that merges the two, and by a hundreds node which merges with this smaller tree by an even higher node). The structure of the syntactic tree is determined by the way a language organizes verbal numbers into groups—e.g., in English and Hebrew numbers are organized into triplets. The tree is not influenced by language-specific order of number-words or irregularities unique to the specific language, which are taken into account at later verbal stages. Importantly, in Dotan and Friedmann's (2018) view, the syntactic tree is a part of the *verbal* representation of the number, rather than a general, abstract representation.

For reading aloud of the number, the constructed hierarchical tree is then *linearized* to the number-word frame by a set of language-specific conversion rules. In this stage, the properties of the specific language and the verbal form of multidigit numbers in it are taken into account—such as the order of number-words and the special rules concerning structure-modifying digits (e.g., 1 in the tens' position), which have an effect on the word frame.

Up to this stage—the number-word frame specifies slots for decimal words (such as the word "thousand" in 42,037) and function words (such as "and"), but still does not specify the number-words themselves (such as "two," "seven"), only their lexical decimal classes. The abstract identity of the number-words is bound into the number-word frame in the next stage, which merges the ordered digits from the numeric buffer into the number-word frame produced in the linearization process. The result of this binding stage is a fully specified, yet abstract (not yet phonologically specified) sequence of words.

In the following stage, the phonological forms of the number-words, the decimal words, and the function words are retrieved from the dedicated phonological storage of number-words (in the phonological output buffer, Dotan and Friedmann, 2015. This buffer is a short-term component that holds ministores of phonemes, morphological affixes, and number-words). This sequence of words undergoes morpho-phonological assembly (the buffer is also responsible for assembling these units), which is then sent to articulation.

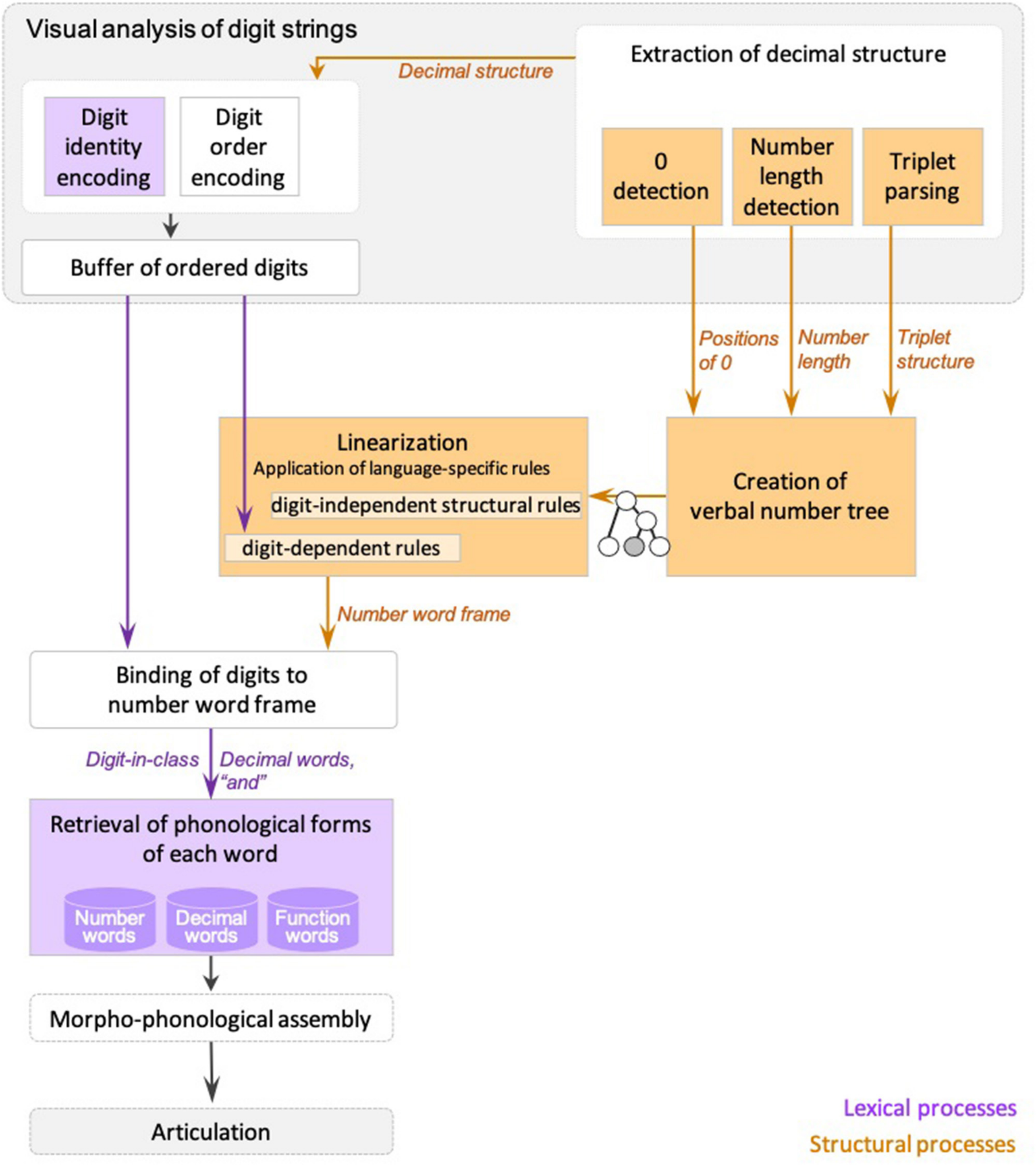


FIGURE 1 | The number reading model from Dotan and Friedmann (2018).

Deficits in each of the components of this multi-staged process, or in the connections between them, lead to various types of dysnumeria (McCloskey et al., 1985, 1986, 1990; Temple, 1989; Noel and Seron, 1993; Cipolotti and Butterworth, 1995; Cipolotti et al., 1995; Cohen et al., 1997; Basso and Beschin, 2000; Deloche and Willmes, 2000; Delazer and Bartha, 2001; Dehaene et al., 2003; Cappelletti et al., 2005; Friedmann et al., 2010; Starrfelt et al., 2010; Starrfelt and Behrmann, 2011; Moura et al., 2013; Dotan and Friedmann, 2015). Next, we will describe various types of dysnumeria that result from impairments to different components and their properties.

1.2. The Dysnumerias: Selective Deficits in Reading Numbers

Dysnumerias in the early stage of numeric-visual analysis—in the extraction of digit-identity, digit-order, or in the extraction of the written number's decimal structure were reported for several individuals, with developmental or acquired impairments.

Digit-order dysnumeria is a deficit in the numeric-visual analysis component that encodes the relative order of digits in a multidigit number. People with this dysnumeria make digit-order errors in reading multidigit numbers, both in tasks that require verbal output and in tasks involving only silent reading, but not in phonological output tasks that do not involve written numbers (e.g., repetition of numbers). Dotan and Friedmann (2018; see also Friedmann et al., 2010) reported on two such women, EY and HZ, who made many digit-order errors in reading aloud and in silent reading but not in tasks that involved number production without numeric-visual input. YS, reported in Friedmann et al. (2010), may also have had this kind of dysnumeria.

Numeric input buffer dysnumeria affects the numeric input buffer without affecting the earlier stages that encode the digits and their order. Such deficit causes digit errors (substitutions, omissions), as well as digit-position errors, and may be susceptible to number-length effect. UN, reported in Dotan and Friedmann (2018) displayed such impairment (in addition to a deficit in number-word frame generation).

A specific deficit may also exist in the ***connection between the numeric input buffer and the later binding stage***, which may cause digit-identity errors in reading Arabic numbers aloud. This may be the deficit of YM, reported by Cohen and Dehaene (1991), and of BAL, reported by Cipolotti et al. (1995). BAL made "lexical" (digit-identity) errors in tasks that required reading Arabic numbers and producing them as spoken number-words. In contrast, his comprehension of Arabic numbers was intact, and he made no errors in the production of numbers written as number-words—suggesting that his deficit was neither in numeric-visual stages nor in phonological output stages. It seems, therefore, that BAL's deficit was in the connection of the written Arabic number to the number-word.

Going back to the early numeric-visual analysis stage, now to deficits in the decimal structure extraction components, ***Number-length dysnumeria*** is a deficit in numeric-visual analysis, selectively affecting the function of encoding the length of the number. Individuals with this impairment make first-digit decimal shifts in multidigit numbers (e.g., reading the number 4320 as 43200). The deficit affects reading aloud as well as silent reading tasks that require processing of number-length (e.g., same-different decision task with numbers differing in length), but does not affect tasks that involve number output with no written input (such as multidigit number repetition). Such patients were reported by Dotan and Friedmann (2018): MA, who had a selective deficit in number length and did not make any other numeric-visual analyzer errors, and HZ (who had zero detection dysnumeria, in addition to her number length dysnumeria). An interesting manipulation used by Dotan and Friedmann, which we used in the current study as well, was the introduction of comma separators to multidigit numbers: when individuals with number-length dysnumeria read numbers with a comma separator, their decimal error rate decreases.

Zero-detection dysnumeria is a deficit in the zero-detector in the numeric-visual analysis stage of exctrating the decimal structure. This dysnumeria causes errors of zero omissions and additions (manifesting as decimal shifts) as well as transpositions of zero with meaningful digits. If this dysnumeria is selective, no decimal shifts are expected in numbers without zero. A patient who has this dynumeria is HZ (Dotan and Friedmann, 2018), who made digit-order errors in numbers including zero (and also, to a lesser degree, in numbers without zero, as she also had a deficit in digit-order).

Decimal shifts may also stem from a deficit in a stage that follows number-length encoding, which uses the number-length information for the creation of the decimal structure of the number. Individuals with ***a deficit in the generation of the decimal structure of written multidigit numbers*** may perform well on tasks that require number-length encoding but still have a selective deficit in parsing the correctly-perceived number-length into triplets, for example, and may fail in the construction of the correct decimal structure of the number. This dysnumeria affects tasks of reading aloud but does not affect tasks that involve phonological output without written-number input such as number repetition. A participant who showed this dysnumeria was ED (Dotan and Friedmann, 2018, 2019). She made many first-digit decimal shifts in number-reading, performed well on visual-analyzer tasks that require processing of length, and benefited from reading the numbers with a comma separator, which did the job of parsing into triplets for her. When she was requested to look at a number without a comma and read it triplet-by-triplet (reading 654321 as "654 and 321"), she still made a similar rate of errors as when she read the number as a whole number, suggesting that she also had a difficulty in parsing into triplets, which might be a part of the decimal structure building. Dotan and Friedmann (2018) ascribed her deficit to a triplet-parsing component in the visual-analysis stage. In the revised model we suggest below, it can be described as a deficit in the conversion of the decimal information from the numeric-visual analyzer to the stage of decimal number frame generation, which may also include information about the division of the number frame into triplets. ED's sister, NL (also reported in Dotan and Friedmann, 2018, 2019), also showed a deficit in this stage (as well as in a later verbal production stage).

Stages of creating the number-word frame may also be susceptible to specific kinds of dysnumeria. The stages that follow the decimal structure extraction in the numeric-visual analysis are the generation of a number frame and the linearization of this number frame onto a verbal sequence. Due to this architecture, people with deficits in these stages show difficulties in reading aloud written numbers, but do not show difficulties in tasks involving the numeric-visual analyzer alone. Dotan and Friedmann (2018) report on several patients who were impaired in these stages: OZ and UN made decimal shifts in tasks involving number production, but not in tasks involving the numeric-visual analyzer alone. They made decimal-shifts both in reading aloud of multidigit numbers, and in tasks requiring production without visual-analysis (e.g., verbal responses to multipliciation and division problems, and UN also made such errors in repetition of multidigit numbers). Such deficit could have resulted from a deficit in the stage of number frame generation and linearization, or in later phonological stages. These shifts occurred mostly in the first digit, rather than in all positions, which excludes the possibility that they resulted from a deficit in phonological retrieval. Dotan and Friedmann (2018) concluded that OZ and UN's deficit is in the ***generation of the verbal syntactic tree or in the stage of linearization into a verbal sequence***.

A somewhat similar type of deficit was the one of ZN, an aphasic patient reported in (Dotan et al., 2014). ZN's deficit manifested itself in reading aloud of two-digit numbers. Dotan and Friedmann (2018) ascribed his deficit, like the deficit of UN and OZ, to either tree generation or linearization.

For our discussion below it is important to remark that Dotan and Friedmann's (2018) model takes both these stages—tree generation and linearization—to be verbal. Because ZN's comprehension of written two-digit numbers was spared—Dotan et al. (2014) and Dotan and Friedmann (2018) concluded that the comprehension of numbers does not require conversion of the multidigit number to its verbal representation, and that comprehension relies on a separate route of building the decimal representation.

However, there might be another approach to these results (suggested by the performance of the case study we report below). It might be that the generation of the number tree (or at least of the decimal number frame) is non-verbal, creating an abstract representation of the decimal structure of the number. Such abstract representation may be part not only of reading aloud but also of number comprehension. In this case, ZN's spared comprehension could be explained by placing his deficit in the verbal linearization stage, rather than in the abstract number frame generation stage.

A different locus, then, can be ascribed to the deficit of NR, reported by Noel and Seron (1993). NR had decimal shifts, most prominently in reading 3-digit numbers, with much fewer errors when only numeric-visual analysis was required (same-different length judgement, and the detection of the number of digits in a number, which can be performed in the visual stage, without comprehension). However, in contrast to ZN, NR's deficit also affected her comprehension of written numbers: she read and understood 3-digit numbers as if they were 4-digit numbers (e.g., 458-> 4508). NR's deficit can be attributed to the abstract number frame generation stage we have just proposed, which is shared by reading and comprehension, or to the connection between the numeric-visual analysis and the number frame generation.

Patient AT, reported by Blanken et al. (1997), made digit-order errors between the tens and the units digits in reading numbers aloud, whenever the German inversion rule (in which units are pronounced before tens) had to apply, but performed flawlessly in comprehending these numbers. It seems, therefore, that he had ***digit-binding dysnumeria***—a deficit at the stage of binding the digits with the decimal frame.

Impairments in later stages of ***the phonological production of number-words*** cause number identity errors and decimal errors (substitution of numbers of different decimal classes) whenever verbal production of numbers is involved, but not in tasks of silent reading that only involve the numeric-visual analysis and possibly also number comprehension. Such cases were reported in Cohen et al. (1997), and for HY and JG in McCloskey et al. (1986). Dotan and Friedmann (2015) also reported on patients who had deficits in number-word production that resulted from a deficit in phonological output stages—a selective deficit in the phonological output buffer alone (SZ, GE) or also in the phonological output lexicon (YL, ZH, RB, and ZC). Impaired production of multidigit numbers due to phonological output deficits were also reported for GS (Girelli and Delazer, 1999; Delazer and Bartha, 2001) and for FA and RA (Marangolo et al., 2004, 2005).

Beyond the phonological output of numbers, some studies also reported selective impairments in writing of number-words. For example, patient BO (Deloche and Willmes, 2000) showed a selective ***impairment in number-word writing***. When presented with written multidigit Arabic numbers, she was able to read them aloud but failed to write them as number-words.

1.3. Why Test Dysnumeria in Sign Language?

These reported cases of specific dysnumerias show that specific stages of the number-reading process may be selectively impaired and cause different patterns of errors. In the current study we examine how a selective dysnumeria manifests itself in a different modality: in sign language. In addition to revealing how different types of dysnumeria may manifest themselves in a different modality, testing number-reading in sign languages has some unique advantages. First, it allows for a minimal comparison between reading of the same multidigit numbers with and without decimal structure: In ISL, multidigit numbers denoting quantity, such as number of objects, are signed with a decimal structure. However, a parallel system of non-decimal numbers exists which is used for signing numbers such as height, age, or bus numbers, and these numbers are signed as a sequence of digits, with no decimal structure (similar to the digital strategy used in some rural sign languages, as reported in Zeshan et al., 2013, and in Lingua dei Segni Italiana [LIS, the Italian sign language] for numbers 21–99, as reported in Mantovan et al., 2019, and as mentioned in Semushina and Mayberry, 2019, for ASL). Such non-decimal numbers (as do decimal numbers) are often signed in ISL in such a way that the digits are signed slightly moving from left to right locations (in right-handed signers), just like the direction in reading and writing written numbers.

Given a deficit in the creation of the decimal structure of multidigit numbers, as we claim our participant has, the existence of the non-decimal system allows for a direct comparison of the participant's signing of the same written multidigit number, once with a decimal structure and once without. Such comparison can provide an important view on the process of multidigit number reading.

Moreover, since sign languages do not have an orthographic system (Mayer, 2017), signers usually learn to read the orthography of the surrounding spoken language. Numbers written as number-words in the surrounding language provide the signer with a number frame, but in a different language and, in fact, a different modality –without providing clues about the verbal number frame in the sign language. This enables testing whether providing a number frame without clues about the required sign-phonology helps in reading numbers aloud. These two unique properties of sign languages can help to disentangle impairments specific to the construction of the number frame from impairments in other stages of the number-reading process.

As we will show below, the case of dysnumeria in sign language on which we report in this study will help in further developing the model suggested by Dotan and Friedmann (2018)—and in expanding it to also include stages required for the comprehension of multidigit numbers and to digit-by-digit reading.

1.4. Signing Numbers in ISL

ISL has a set of handshapes representing digits, as shown on **Figure 2**. In multidigit numbers in ISL, the decimal class of all but the unit digits is morphologically marked: the digit handshape is incorporated with the movement representing the decimal class, which together create the number-sign for this class. For example, as presented in **Figure 3**—in the ISL sign TWENTY, the handshape representing the digit 2 is incorporated with the movement representing tens, and in the sign TWO-HUNDRED, the handshape representing the digit 2 is incorporated with the movement representing hundreds.

As in many other languages, such as English and Hebrew, multidigit numbers in ISL are organized into triplets—for example, the number 985672 is signed as NINE-HUNDRED EIGHTY FIVE THOUSAND, SIX-HUNDRED SEVENTY TWO. An example of a multidigit number in ISL can be seen in **Figure 4**.

ISL also has a unique structure for teens, but unlike many other languages (e.g., English, Hebrew), the part denoting teens is signed prior to the part denoting the specific units digit (i.e., TEEN-FOUR), so the tens and the units are still signed in the order in which they are written in the Arabic number (see **Figure 4**).

As can be seen on **Figure 2**, some pairs of digit signs in ISL (i.e., 8–3, 9–4) differ only in the use of the non-dominant hand. ISL users sometimes omit the non-dominant hand during signing, and use mouthing (moving the lips according to the parallel Hebrew word) to disambiguate the sign (e.g., 8 can either be signed with two hands- the dominant hand signing THREE, the non-dominant hand held with fingers spread, palm to the interlocutor, or it can be signed only with the dominant hand signing THREE, and the lips articulating "shmone," the Hebrew word for eight). This phenomenon of non-dominant hand omission was probably boosted by the use of cellphones which are held in one hand, leaving only one hand for signing (during video calls, but also simply when the hand is occupied with holding the phone). There are also cases of hand addition, where 3 is signed like 8, which mainly occurs in the context of a neighboring two-handed number sign in which the non-dominant hand is kept in the "5" shape.

2. THE PARTICIPANT

Nomi is a 45-year-old woman, who is congenitally deaf and uses sign language as her main means of communication. As a daughter of hearing parents, she did not use sign language from

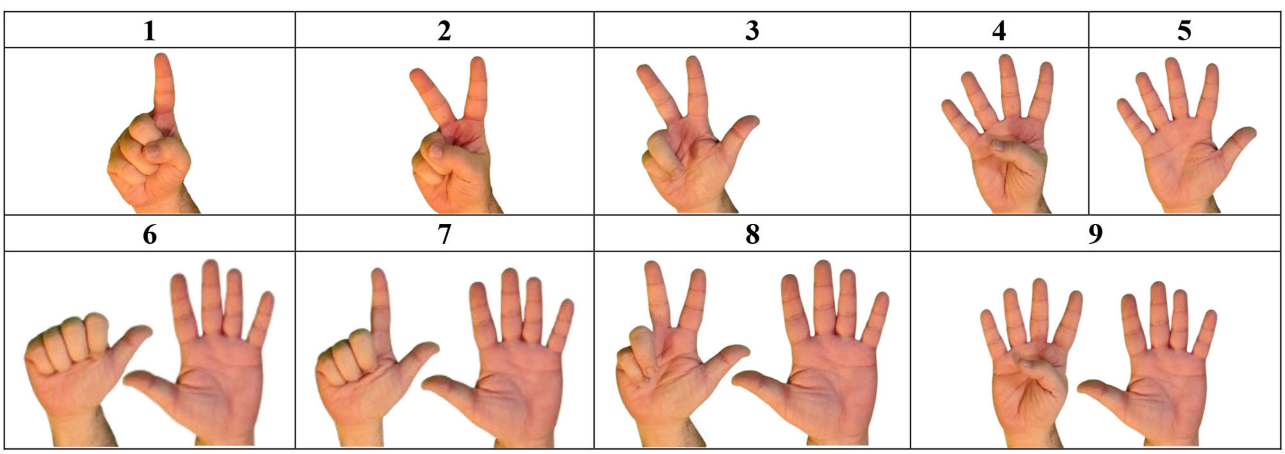


FIGURE 2 | Numbers 1–9 in ISL illustrating phonological similarities between number-signs (e.g., the similarity between 3-8 and 4-9. See **Appendix** for variants of some of the number-signs).

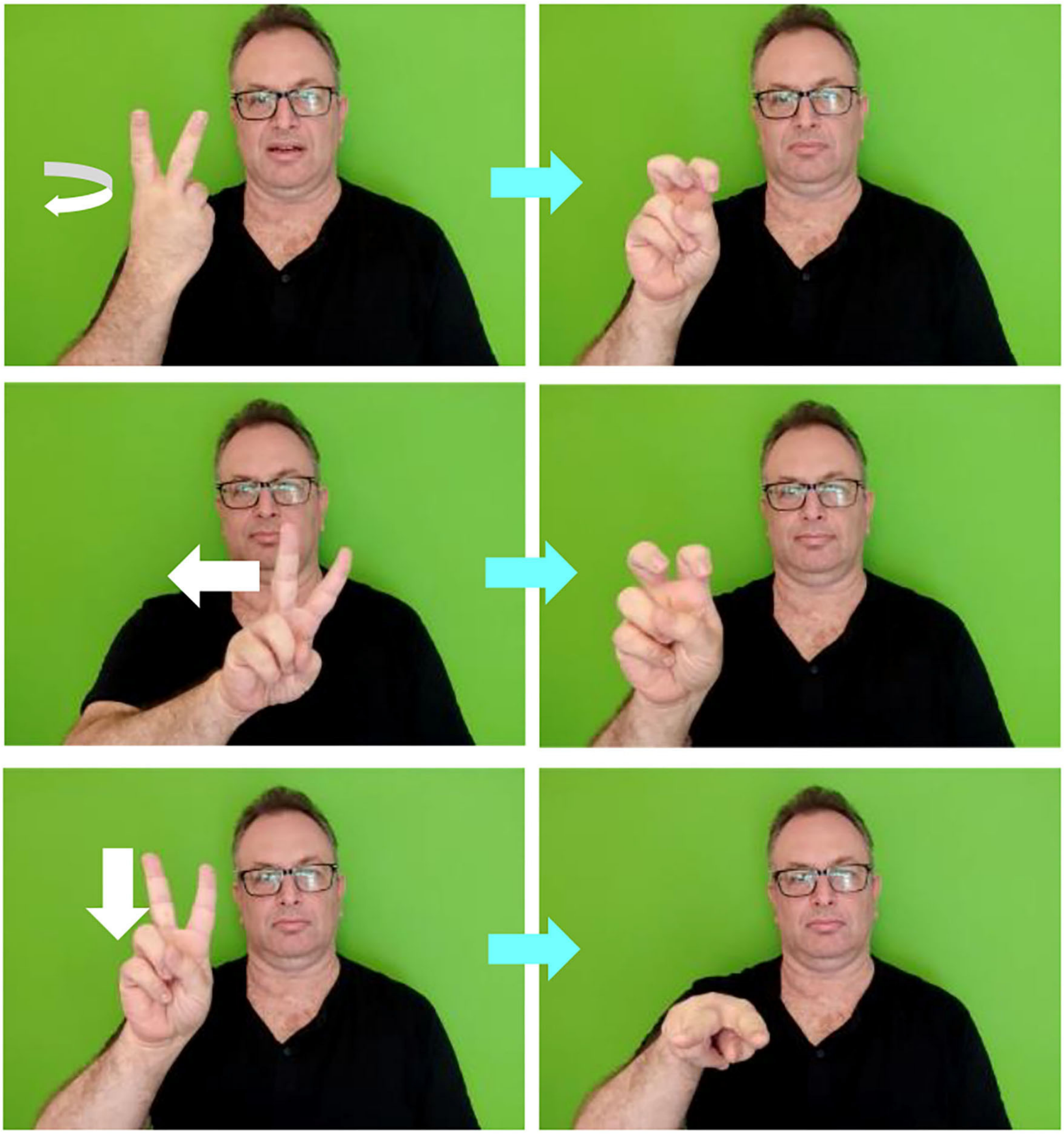


FIGURE 3 | Decimal number-signs in ISL. Top: TWENTY, middle: TWO-HUNDRED, bottom: TWO-THOUSAND. (To see the videos, click the arrows).

birth, but was occasionally exposed to ISL from her deaf signing grandparents. She started to sign consistently only at a later age, and since she was 16, ISL is her main means of communication. Nomi had been using hearing devices since she was 1;6-year-old until the implementation of her cochlear implant around the age of 30 years, which she is using consistently when she communicates with hearing people since then.

Nomi told us that she always felt difficulty with numbers—and at school she was diagnosed as having a "learning disability," with no further specification of her exact deficit.

We assessed Nomi's conceptual abilities using an odd-one-out task, in which she was presented with 30 sets of 4 pictures and was requested to select the one that is not related to the other pictures (e.g., three wild animals and a dog; three fruit and a vegetable, MILO test, Friedmann, 2017). Nomi performed perfectly (100% correct) on this task—indicating that her conceptual abilities are intact.

We also assessed her lexical knowledge in ISL, as well as her lexical retrieval components: the semantic lexicon, the phonological output lexicon, the phonological output buffer, and the connections between them, using a picture-naming task (SEMESH, Biran and Friedmann, 2004). In the picture naming task she was presented with 91 pictures and was requested to sign their names. Here again, Nomi performed flawlessly, with

FIGURE 4 | The multidigit number 14752 in ISL (To see the videos, click the pictures).

100% correct responses– indicating intact lexical retrieval in ISL, including intact phonological output components, and rich lexical knowledge.

She also performed not differently from other ISL signers on a serial recall task of ISL digits (adapted from the FriGvi test battery, Friedmann and Gvion, 2002; Gvion and Friedmann, 2008), and even in the higher range of the control performance in serial recall of ISL signs (long sign span test from the SIMBA test battery, Haluts and Friedmann, 2019). Her digit span was 4 (performance of controls—signers without WM deficit: $n = 10$, mean = 5.15, SD = 0.85) and her long sign span was 4.5 (controls: $n = 17$, mean = 4.4, SD = 0.6).

Her performance in the serial recall task indicated that her phonological working memory abilities, including the phonological input and output buffers, are intact, and the long-sign span also confirmed that her lexical knowledge of ISL is comparable to other ISL signers, as also indicated by her performance in the naming task and her word-reading-then signing task reported below.

3. GENERAL METHOD

We administered the tests detailed below to Nomi in six video (Zoom) sessions, each session was 1–2 h long. The sessions were video-recorded and analyzed and scored by all three authors separately. Nomi signed an informed consent prior to participation and was paid 50 ILS per hour. She was informed that she can stop her participation at any time and could take as many breaks as she wanted during the sessions. The research was approved by the Tel Aviv University Ethics Committee. All comparisons between Nomi's performance and the control groups were performed using Crawford and Howell (1998) *t*-test for the comparison of an individual with a control group, and the comparisons between Nomi's performance on the different tasks were done using McNemar exact or Chi-squared tests.

3.1. Control Group

Nomi's performance on the tests was compared to a control group of 10 deaf adults who use ISL as their main means of communication. They were tested in the same settings and procedure as Nomi did, in virtual Zoom sessions. Like Nomi, all of them have hearing parents and therefore did not acquire ISL from birth, but rather at a later age. They were 6 women and 4 men, aged 25–48 (mean = 32, SD = 6); nine of them were deaf from birth, and one gradually lost her hearing from birth until age 3. In the word-reading tests we also compared Nomi to control groups of typically-hearing native Hebrew speakers.

3.2. Error Analysis

Error analysis was conducted in the following way:

Substitutions of one digit with another (e.g., 589 → 579) were coded as *identity errors*; transpositions of digits that appeared in the number but were signed in an incorrect order (e.g., 3527 → 3257) and any production of a digit that appeared elsewhere in the number (985674 was signed as 96...) were coded as *order errors;*

Varieties that are used sometimes by ISL signers (and appeared also in our control group's signing), were not counted as errors. Therefore, additions and omissions of the non-dominant hand held in the "5" shape in signing a number were coded as acceptable. *Phonological errors* were other errors (not digit identity or number of hands errors) that involved the location, movement, or handshape of the digit (in effect, these errors were extremely rare, with a total of 3 for Nomi and for the control group together).

Errors in the decimal structure of the number (often regarded as "syntactic errors") were coded as *decimal errors*. For example, had a participant started reading the number 45638 as 4563..., or made a decimal error resulting in an ungrammatical production (e.g., 'forty fifty thousand...', in which the decimal-class error was in the middle of the number), the error was coded as a decimal error. When the error resulted in shifting the decimal position of one or more digit, resulting in a decimal frame of an incorrect size, it was coded as a specific type of decimal error—*a decimal shift* (e.g., 80001 → 8001. Errors were coded as decimal shifts also when they were self-corrected at some point during the reading of the number 86952 → eight thousand, six hundred... ✓)[1]. We classified the decimal shifts according to the shift direction (whether it resulted in a smaller or a larger decimal class). Another type of error that was coded as decimal (and was much less frequent than shifts) was "thousand" omission—omitting the word/sign "thousand" between the triplets, which may suggest production of two separate triplets (with two smaller decimal frames) instead of the full number.

Finally, numbers for which the participant refused to sign and asked to move to another item were coded as "didn't sign" errors.

4. A DEFICIT IN MULTIDIGIT NUMBER READING AND ITS SOURCE

4.1. The Deficit: Reading Written Multidigit Numbers "Aloud"- Reading and Then Signing

Method

Nomi was presented with 60 multidigit numbers written one above the other (MAYIM battery, Dotan and Friedmann, 2014). Of these, 30 were shorter numbers (10 three-digit numbers and 20 four-digit numbers), and 30 were longer numbers (25 five-digit numbers and 5 six-digit numbers). The numbers of the various lengths were randomly ordered; 33 of the numbers did not include zero, and 27 included a single zero-digit. The task was similar to "reading aloud" task in spoken language: Nomi was asked to read each number and then sign it in ISL. We will henceforth relate to this test as "the baseline".

Results

Nomi made errors on 18 of the numbers she read in this task (28%), and her performance was significantly lower than the controls' (mean correct = 91%, SD = 6%, Crawford and Howell's $t(9) = 3.34$ $p = 0.004$). Nomi's most pronounced error type was decimal errors, which occurred only in 5–6 digit numbers: she made 11 decimal errors (37% of the 5-and 6-digit numbers)—significantly more than the controls (mean = 6%, SD = 4%, Crawford and Howell's $t(9) = 2.77$, $p = 0.01$). For instance, when reading the number 89712 she signed "EIGHT-THOUSAND... NINE... NINE... NINE...," and then corrected herself; When reading the number 985723 she signed "NINETY-EIGHT THOUSAND..., FIVE THOUSAND SEVEN HUNDRED TWENTY-EIGHT," making a decimal as well as a digit-identity error. Most of Nomi's decimal errors were in the first, leftmost, digits, and the direction was always toward a smaller decimal position (89712 → 'eight thousand...' but not 'eight-hundred ninety-seven thousand...'). She could not read correctly any of the 6-digit numbers, on which she either made a decimal error (on 3 of the 5 numbers in this length) or said that she cannot sign the number and asked to move on to the next one (2 numbers). She also made 3 non-decimal errors– 2 digit identity errors and 1 phonological error (which she immediately self-corrected). During the test Nomi reported that reading the long numbers is very difficult for her, and that she doesn't know how to sign them.

The next step was to try and find the origin of her deficit in reading multidigit numbers. Decimal errors in reading can arise from a deficit in the numeric-visual analysis of the written-number input, from a deficit in the conversion of written input-to the phonological (signed) output, or from a deficit in the phonological output processes.

4.2. Better Reading of the Same Multidigit Numbers as Non-decimal Numbers

To examine the source of Nomi's deficit in reading multidigit numbers, and specifically, to examine her numeric-visual analysis, we used a special property of ISL: some multidigit numbers that represent quantity (such as 61 seashells, or 123 new students) are signed as decimal numbers, with decimal morphology and structure, similar to numbers in spoken languages (e.g., 123 [students] is signed ONE-HUNDRED TWO-TENS THREE [the sign for 1 with hundred-morphology, the sign for 2 with tens-morphology, and the sign for 3]). However, a parallel system of multidigit numbers exists, which is used mainly for numbers that do not symbolize quantity (e.g., social-security numbers, bus numbers) and for certain measurement units (e.g., height, weight, age), in which the numbers are signed as a sequence of digits without a decimal structure, and without decimal morphology. For instance, when signing the (old) age of 123, the number will be signed "ONE TWO THREE," with no decimal structure.

This allowed us to isolate the conversion to a decimal structure from the other components of multidigit number reading: visual-analysis of the sequence of digits and phonological production of the sequence of number-signs, and compare Nomi's signing

[1] Zero-position errors, e.g., 8030 → 8003, which occurred 4 times in the control's reading (and never in Nomi's reading) were coded separately from decimal shifts.

of the exact same multidigit numbers with and without the conversion to a decimal template.

Method

Nomi was presented with 60 numbers, same numbers as in the baseline task (section 4.1), presented in the same way. The only difference was the instruction: now she was told it is a list of passwords, and she was again asked to read and sign.

Results

Once she was requested to sign the numbers without their decimal structure, Nomi was far more accurate. She commented that signing numbers this way was much easier for her than signing them as quantity numbers. She signed with confidence, and, like the control group- from left to right. She made only 2 digit-order errors (transpositions of adjacent digits), so her performance was significantly better than her reading of the exact same numbers with a decimal structure (97% correct), McNemar test $p = 0.0001$ (one-tailed).

Nomi's performance in this test was comparable to that of the controls' (mean = 98%, SD = 2%, Crawford and Howell's $t(9) = 0.80$, $p = 0.22$).

Her far-better performance on reading the same numbers when she did not need to convert them to decimal structures indicates that her visual-input processes themselves are intact and cannot be the source of her deficit. Had she had a deficit in the perception of number-length or in the position of zeros, we would have expected similar errors in reading digit-by-digit—errors of omission or doubling of digits, or errors in the position of zero, which she did not show. This finding, suggesting preserved visual analysis of number-length, together with the decimal errors she made in reading the same numbers decimally, hint at a deficit related to the decimal structure, in a stage later than the numeric-visual stage.

4.3. Presentation of 6-Digit Numbers With an Instruction to Read Each Triplet Separately

If Nomi's impairment is indeed in converting number-length of long numbers into number frames, another manipulation expected to help her in reading long numbers is reading the number triplet-by-triplet, breaking the long number into two shorter 3-digit decimal numbers (see Dotan and Friedmann, 2019, for this manipulation in dysnumeria).

Method

Nomi was presented with 20 6-digit numbers (all with 6 unique digits) and was requested to split the numbers in two triplets such that she read the first 3 digits as one decimal triplet, and the last 3 digits as another decimal triplet (e.g., when presented with 123456, she was expected to sign ONE-HUNDRED TWENTY-THREE; FOUR-HUNDRED FIFTY-SIX). Here we provided her both with the length of the number (as there were 6 digits in all numbers presented in this task), and exempted her from the need to sign numbers >999, namely, she did not need to construct a tree higher than hundreds and did not need to produce thousand, 10-thousand, or 100-thousand number-signs.

Results

Nomi did not make any errors in this test and signed correctly all of the triplets (100%). She also used mouthing of the Hebrew decimal number correctly in all of her productions. This suggests, again, that when she does not need to create a number frame higher than 3-digits, Nomi shows no difficulties. Her performance in this task thus further supports the idea that her deficit arises when she needs to create a decimal number frame for numbers longer than 3-digits.

4.4. Additional Evidence for Intact Numeric-Visual Input Processes: Three Same-Different Tasks and a Sequence Decision Task

Nomi's decimal errors in reading multidigit numbers could have arisen from a deficit in the input stage of numeric-visual analysis, in processing number-length or digit position; alternatively, it could stem from a deficit in creating a number frame that would match the written number. Her good performance in the digit-by-digit and the triplet-by-triplet reading of multidigit numbers, described above in sections 4.2 and 4.3, indicates that she did not have a digit-order deficit in the numeric-visual input stage (nor did she have a deficit in digit-identity, but this component anyway is not a candidate for decimal errors). It is a bit trickier to guess how a number-length deficit would manifest itself in digit-by-digit reading: it could cause digit-omission, zero insertion or omission, and perhaps digit-duplication. Nomi did not make such errors in the digit-by-digit task and the triplet-by-triplet task. To further examine Nomi's visual input processes of number-reading, and to examine directly her numeric-visual encoding of number length and order, we tested her in four tasks that involve the numeric-visual analysis of written numbers, without requiring production.

4.4.1. Same-Different Tasks

Since same-different decision tasks of written Arabic numbers involve only numeric-visual input, and do not require any verbal output, nor do they require the creation of a number frame, they allow for a direct and specific examination of the numeric-visual analyzer. In the same-different tasks we examined the encoding of number-length, digit-identity, and digit-order, by way of manipulating the differences between the two numbers—they could either differ in length, identity, or order.

To manipulate number-length without the concomitant manipulation of identity and order, we used the duplication of digits (e.g., deciding whether 9939 and 99399 are the same). This allowed for the specific and direct examination of length extraction in the numeric-visual analyzer, since the numbers differing in length do not also involve a change in digit-identity (they contain the same unique digits), neither do they involve a change in relative digit-order (the digits are written in the same order, with an additional duplicate in one of the number's ends). This allowed us to examine whether Nomi's decimal errors result from a deficit in length extraction in the numeric-visual analyzer.

4.4.1.1. Same-Different Decision: Pairs of Numbers Presented Together

Nomi's ability to detect differences between written numbers was first tested in a same-different task in which the pairs of numbers were presented on a screen together, one above the other, without time limitation. Nomi was requested to judge, for each pair, whether the numbers were identical or different. The task included a total of 118 number-pairs, all including a repeating digit and another digit, presented in two sessions: 27 pairs of 3–6-digit numbers that differed in length (created by the addition of another instance of the repeating digit, e.g., 99399 and 9939), 20 pairs that differed in order (e.g., 99899 and 98999), 16 pairs that differed in number-identity (e.g., 97999 and 95999), and 55 identical pairs.

4.4.1.2. Same-Different Decision: Numbers Presented Sequentially

To rule out the possibility that Nomi succeeded in the first same-different task because the comparison of simultaneously presented pairs was too easy, we also administered a similar task with delayed comparison, where the numbers of each pair were presented on a computer screen one after the other (using Testable). The first number of the pair was presented for 1200 ms, followed by a 500 ms masking period, and then the second number appeared for 1200 ms. Then a question mark appeared on the screen and the participant decided whether the numbers in the pair were the same (by pressing the "l" key) or different (by pressing "a"). This task included a total of 40 pairs: 10 pairs of 3–6 digits numbers that differed in length (7 of which were pairs of 4–5- and 5–6-digit numbers), 8 pairs that differed in number identity, and 22 identical pairs.

4.4.1.3. Number Matching (99499)

Another way in which we examined Nomi's number-reading when no verbal production is required was a matching task. Nomi was presented with a page on the screen, on top of this page appeared a reference target multidigit number and 36 numbers printed underneath. She was requested to circle all numbers that were identical with the reference number, as accurately and quickly as possible.

The reference number was 99499, and the 36 numbers beneath it included 5 items that differed from the reference in digit-order (by transposing 9 and 4), 25 that differed from it in length (by adding or subtracting instances of the digit 9), and 6 items that were identical to it.

4.4.1.4. Same-Different Tasks Results

Nomi showed very high performance on all three same-different tasks, as summarized in **Table 1**. When presented with the pairs written one above the other (4.4.1.1), she made no number-length, digit-order, or digit-identity errors at all, and had only 2 misses of identical pairs that she did not mark (4% of the identical pairs).

In the sequentially-presented numbers task (4.4.1.2), she made no number-length or digit-identity errors and had only one miss (pressing "different" for a same pair, 98% correct). This performance is not different from that of the control group (mean = 95%, SD = 9%, Crawford and Howell's $t(9) = 0.33$, $p = 0.38$).

In the number matching task (4.4.1.3), Nomi made no length or order errors– she never marked a number that was different from the reference number. She did miss 2 of the 6 identical numbers.

In all same-different tasks, thus, she never made any errors of mistakenly marking a different pair as similar, and importantly—she never mistook a pair of numbers *differing in length* as a similar pair.

In addition, Nomi explicitly reported that these same-different tasks were easy for her, she even used the sign for "fun."

4.4.2. Additional Evidence for Intact Numeric-Visual Input Processes: Sequence Decision

Method

Nomi was presented with 118 4-digit numbers (sequence decision task, MAYIM battery, Dotan and Friedmann, 2014), 60 of which were strictly monotonic increasing sequences of consecutive digits (e.g., 1234), 33 included a transposition between two adjacent digits, such that the serial sequence was violated (e.g., 1243), 25 included a substitution of one of the digits such that it did not create a monotonic increasing sequence (e.g., 1274). Nomi was asked to mark all the numbers in which the digits created a consecutive monotonic increasing sequence.

Results

In the sequence decision task, Nomi made only 2 errors (<2%)—both on the same sequence (5687) which she marked even though it was a transposed sequence. This indicated that her digit-order encoding is intact.

4.4.3. Interim Summary: Assessment of Nomi's Numeric-Visual Analysis

Nomi's far better reading of multidigit numbers when she did not need to convert them onto decimal structures larger than a triplet, as well as her good performance on the same-different and sequence decision tasks all point together to the same conclusion: her numeric-visual input processes are intact and cannot be the source of her deficit. She did not have a deficit in number-length perception, nor in digit-order or in zero-position, which could be the basis for her decimal errors in reading-then-signing multidigit numbers. Her flawless performance in reading the same multidigit numbers digit-by digit indicates that her numeric-visual analysis of digit identity and order is intact. The locus of her deficit, then, has to be a later stage in the number processing model: either in the conversion of the written number into its decimal frame, in constructing the decimal structure of the number frame, or in later phonological output stages.

4.5. The Decimal Errors Do Not Stem From a Phonological Output Deficit

4.5.1. Multidigit Numbers Written in Number Words

Had Nomi's deficit been in the number production stages in the phonological output buffer, responsible for retrieval and assembly of number-words, we would expect her to make errors

TABLE 1 | The number tests and Nomi's performance in them.

Test (Section)	Description	# items	% correct	Decimal errors	Other errors
Reading and signing (RS) multidigits-the baseline task (4.1)	Read-then-sign multidigit numbers written as Arabic numerals	60	70%	11 decimal errors in 5–6 digit numbers, 2 "don't know" errors in 6-digits.	2 identity errors and 1 phonological error (which was immediately self-corrected)
RS multidigits as non-decimal numbers (4.2)	Read-then-sign same numbers as in the baseline task, but this time digit-by-digit	59	90%	–	2 serial order errors with immediate corrections
RS Multidigits—Triplet-By-Triplet (4.3)	6-digit numbers written as Arabic numerals triplet-by-triplet	19	100%	No decimal errors	No errors
Same-different multidigits: simultaneous (length, order, identity) (4.4.1.1)	Pairs of multidigit Arabic numerals presented simultaneously, asked to judge same/different	118 pairs	98%	–	2 misses of identical pairs
Same-different multidigits: sequential (length, identity) (4.4.1.2)	Pairs of multidigit numbers presented one after the other with masking between them, asked to judge same/different	40 pairs	98%	–	1 miss of an identical pair
Same-different multidigit matching (Length, order) (4.4.1.3)	Target multidigit number and 36 numbers below it, asked to mark numbers similar to target	36	94%	–	2 misses of matching numbers
Sequence decision in multidigits (4.4.2)	4-digit numbers written as Arabic numerals, mark the ones that contain a sequence of monotonic increasing consecutive digits	118	98%	–	2 order errors
RS multidigits written as number words (4.5.1)	Read-then-sign multidigit numbers written in Hebrew words	18	94%	No decimal errors	1 whole-unit error
Repetition of multidigits (4.5.2)	Repetition of multidigit numbers presented as ISL signs	45	42%	2 decimal errors	Many identity and order errors
Repetition of multidigits with fewer number-words (multiple zeroes) (4.5.3)	Same task as (4.5.2), with numbers requiring fewer number-signs	37	84%	No decimal errors	6 errors—4 identity errors, 1 order error, and 1 morphological error
RS multidigits with fewer significant digits (multiple zeroes) (4.5.4)	Read-then-sign same numbers as (4.5.3) written as Arabic numerals	40	60%	9 decimal errors (23%), and 6 "don't know" responses	1 digit identity error (immediate self-correction)
>5500—written input no comma (4.6.1)	Decide whether multidigit Arabic numerals are >5500	20	80%	4 errors: all 5-digit numbers in which the first digit <5	-
>5500—signed input (4.6.2)	Decide whether signed multidigit numbers are >5500	30	97%	1 error –first item	-
Multidigit number-comparison (4.6.3)	Decide which of a pair of multidigit numbers written as Arabic numerals is greater	68 pairs	97%	3 Errors in longer numbers. Much longer RTs than controls in same-length 6-digit pairs, and in different-length pairs with incompatible first digit.	-
RS multidigits with a comma separator (4.7.1)	Read-then-sign same numbers as in the baseline task (4.1), but with a comma separator	60	97%	2 decimal errors	–
>5500—written input with comma (4.7.2)	Same task as in written > 5500 (4.6.1), but with a comma separator	30	100%	–	No errors
Numeral incorporation (NI) (4.5.5)	Read-then-sign (translate to ISL) written Hebrew sentences containing NI phrases	42	100% on relevant structures		No errors in the NI structures

Dark gray, significantly more decimal errors than the controls; Light gray, significantly more than the controls but much fewer than in the reading and signing baseline task.

not only when she reads multidigit numbers, but also when she produces them without reading. To test Nomi's production of multidigit numbers without reading of Arabic numerals, and in a way that provides her with the decimal structure of the numbers, we took advantage of her being Hebrew-ISL bilingual, and provided her with the decimal structure of the numbers in a separate system- a written spoken language, which does not provide her with the needed number-signs in ISL. We presented her with multidigit numbers written in Hebrew number-words (e.g., five hundred twenty-four; Word-To-Number test, MAYIM battery, Dotan and Friedmann, 2014). Notably, number signs do not show one-to-one correspondence to written number words (for example, "four hundred" is written as two Hebrew words but corresponds to a single number-sign). This allowed us to test Nomi's production of multidigit numbers once the decimal structure is provided, and thus tease apart the conversion of the written number to a decimal structure, and the production of the decimal structure of multidigit numbers.

Method

The test included 18 multidigit numbers written in Hebrew number-words. The target signed numbers were composed of 2, 3, 4, 5, and 6 number-signs (4, 5, 3, 5, and 1 items, respectively). Nomi was requested to read the Hebrew written words and, when she finished reading the whole number, sign the number in ISL.

Results

Nomi made no decimal errors at all in this test. She only made one whole-unit error (in response to the Hebrew number-words "three thousand and-nine teen" she signed "3900" (parallel to "three thousand and-nine hundreds"), and then corrected herself. This performance indicates that her deficit in the decimal structure of multidigit numbers does not stem from impaired retrieval of the correct number signs in the correct decimal morphology at the stage of the phonological output buffer, nor from a deficit in holding these signs and assembling them in this buffer, but rather from a component that converts the written Arabic numbers into the number frame.

4.5.2. Repetition of Multidigit Signed Numbers

Another way to examine Nomi's production of multidigit numbers without involving written input of Arabic numerals was to ask her to repeat a list of numbers. These were the same numbers she had read on the baseline reading task (4.1), list on which she made many decimal errors.

Method

The task included 45 of the multidigit numbers from the baseline number reading task. The signed numbers included 3-, 4-, 5-, and 6-digits/ number signs (9, 12, 20, and 4 items per length, respectively). A native signer signed each multidigit number and Nomi was requested to repeat it immediately, as accurately as possible.

Results

It seems that the 4–6-digit numbers exceeded Nomi's signed digit span (which was 4, see Participant description) so she omitted and substituted digits in 75% of the 5–6 digit numbers she tried to repeat and in 50% of the 4-digit numbers, with a general percentage correct of 42%. However, crucially, her pattern of errors was completely different from the one she displayed in number-reading: she made only 2 decimal shifts in her repetition, both with correct Hebrew mouthing despite the manual error, significantly fewer decimal shifts than in reading the same numbers (McNemar test $p = 0.02$). She also had repetition-errors on shorter numbers than in reading: in 4- and even 3-digit numbers.

4.5.3. Repetition of Signed Multidigit Numbers With Fewer Unique Digits and Fewer Number-Signs

Method

Nomi's high rate of non-decimal errors in repetition of multidigit numbers in the previous task might stem from exceeding her working memory capacity. We therefore tested her repetition of multidigit numbers in a task that was less taxing for her phonological working memory, which used multidigit numbers (3–6 digits) with fewer (3–4) significant digits and fewer number-signs (2–3). For example, 403000 is a 6-digit number, the same length as the longest numbers in the baseline list (4.1) section and Its Source, but it requires only 3 number-signs in production (four-hundred, three, thousand). This test included 7 three-digits numbers, 9 four-digit numbers, 18 five-digit numbers, and 6 six-digit numbers.

Results

In repetition of these numbers Nomi made no decimal errors at all. She did make 4 number-identity errors (e.g., substitutions of one number-sign with another), one order error (i.e., she repeated 736 as 763) and one error of number morphology (in repetition of the number 3021, she signed separate signs for "three" and "thousand" instead of the ISL single sign "three-thousand"). She reported at the end of the test that she was not focused, so this may be the cause of these errors. Importantly, here again the errors were not unique to longer numbers—half of them happened with 3–4-digit numbers, and, crucially, were not decimal errors.

4.5.4. Reading and Signing Long Multidigit Numbers With Multiple Zeros

The last three experiments demonstrated that Nomi has no decimal structure problems in producing multidigit numbers once no written Arabic numerals are involved. In the next task we took an additional view as to the question of whether her decimal errors in number-reading (in the baseline task) stemmed from a phonological overload of number-signs. To examine this, we tested her reading of multidigit numbers of the same length as in the baseline task, but this time with fewer number-signs.

Method

We asked Nomi to read 40 multidigit numbers with fewer significant digits (3–4) and fewer number-words (2–3). The numbers included multiple zeroes instead, and were the same numbers used in the task reported in the previous section (4.5.3).

FIGURE 5 | Numeral incorporation in ISL. Top: TWICE, middle: TWO-MONTHS, bottom: TWO-HOURS (To see the videos, click the arrow in each of the picture pairs).

Results

Even though these numbers required far fewer number-words than did the numbers in the baseline list, Nomi still made many errors in this task (43%) and showed a similar error rate (χ^2 =1.65, $p = 0.20$) to the one she showed in reading-then-signing numbers without multiple zeros (in the baseline task). This result indicates that her errors did not result from an overload of number-signs in the phonological output buffer, but rather have a different origin.

Her error pattern in this test was also similar to the one she showed in the baseline task: She made 9 decimal errors (23%), and refused to sign 6 numbers, remarking these were too long for her. Like in the baseline, the majority of her errors occurred in the longer numbers: she could not sign any of the 6-digit numbers (one decimal error and 5 refusals), and could sign only 56% of the 5-digit numbers (7 decimal errors and one refusal). She made only one decimal error in a 4-digit number, and no errors in the 3-digit numbers. Here again, as in the baseline task, the direction of decimal shifts was almost exclusively toward the smaller decimal position.

Nomi's performance in this test was significantly worse than her performance in the parallel repetition task (in the previous section), in which she was required to repeat exactly the same 40 numbers (McNemar test $p = 0.01$).

Just like when reading and signing the numbers without multiple zeros (in the baseline task), Nomi struggled in this test. She said that it was very difficult for her, and that she found the long numbers especially hard.

4.5.5. Numeral Incorporation

Nomi's decimal errors were mostly decimal shifts, in which she produced an incorrect decimal class for the correct digit. In ISL, the decimal classes are marked morphologically on the number. Therefore, we wanted to rule out the possibility that the decimal errors originated in a deficit in morphological incorporation of numbers. ISL allows for a direct assessment of this question. As in other sign languages (Liddell and Johnson, 1989; Taub, 2001; Fuentes and Tolchinsky, 2004; Meir and Sandler, 2007; Fischer et al., 2011; Semushina and Mayberry, 2019), ISL uses morphologically complex structures in which a number-sign is incorporated into a base morpheme, usually denoting a time expression or a pronoun—to create a single morphologically-complex sign (e.g., EIGHT-YEARS is a single ISL sign made from the handshape of the number 8 and the movement and location of the sign YEARS, and THREE-OF-US is a sign made from the handshape of the number 3 and the movement and location of the sign US, see **Figure 5** for some examples). If the decimal errors emerged from a morphological difficulty in numbers incorporated in morphologically-complex structures, these non-decimal morphological structures should be affected as well. Additionally, morphological incorporation takes place in the phonological output buffer (Haluts and Friedmann, 2020), so these constructions could also serve as another assessment of Nomi's phonological output buffer.

Method

To test Nomi's production of structures with numeral incorporation, we presented her with 42 written Hebrew sentences containing numbers—19 of which contained structures signed as numeral incorporation in ISL, and asked her to translate them into ISL.

Results

Nomi made no errors at all in the structures involving numeral incorporation. Her good performance on these structures, then, rules out a deficit in the complex morphology of numbers as the basis of her decimal errors, and points to a difficulty that is unique to multidigit numbers.

Additionally, Haluts and Friedmann (2020) showed that signers with impairments to the phonological output buffer make whole-unit errors in these morphologically-complex structures. The finding that Nomi performed well on these structures, as well as her consistent direction of errors - always toward a number which is smaller in one decimal-position, provides another support for the conclusion that her phonological output buffer is intact.

4.5.6. Interim Summary: Assessment of Nomi's Phonological Output

The above tasks indicated that Nomi's decimal errors did not result from a deficit in phonological output processes of selecting the correct number signs including their correct decimal morphology, holding them, and assembling them into a whole multidigit number. She did not make any decimal errors in signing multidigit numbers written as Hebrew words, even though this task requires the same stages of phonological selection, holding, and assembly of complex number-signs. In repetition of multidigit numbers she made almost no decimal errors. When reading long multidigit numbers that require fewer number-signs in production (which reduces the load on the phonological output buffer) she still made the same rate of decimal errors. Finally, she showed normal production of numeral-incorporation structures, which shows that she can produce morphologically complex numerical structures. As we will show below, the addition of a comma-separator, a visual manipulation that does not affect phonological output, significantly reduced her decimal errors. Together with her intact numeric-visual analysis, these results point to a deficit in a decimal structure stage that follows visual input and precedes phonological output.

4.6. Same Impairment in the Comprehension of Written Multidigit Numbers

In the previous sections we have seen that Nomi makes decimal errors in reading and then signing multidigit numbers, and that this deficit cannot stem from her numeric-visual analysis or from the phonological output stages, which were intact. We suggested that the deficit is related to the creation of the decimal number frame for written multidigit numbers. The question now is whether the deficit only affects decimal structure required for phonological production, or whether the deficit also affects tasks that do not require phonological output, such as comprehension tasks.

4.6.1. Impaired Comprehension of Written Numbers: > 5500

Method

To test Nomi's comprehension of multidigit numbers (in a test that does not involve production), we presented her with 20 multidigit numbers printed on one page, and asked her to mark all numbers that are greater than 5500. The reference number (5500) was given to her in signing. The multidigit numbers included numbers of different lengths (3–5 digits, 5 of them with 5 digits) without a comma separator, randomly scattered on the page. Importantly, the five 5-digit numbers, which, obviously, were all greater than 5500, included four numbers for which the first, leftmost digit was smaller than 5. If her deficit in decimal structure affects comprehension of written numbers in the same way it affects reading-then-signing, we would expect that she might apply to these numbers a 4-digit-number structure instead, and then may understand them as smaller than 5500.

Results

Of the five 5-digit numbers, Nomi marked only the one in which the first digit was >5 (93061), but missed all other four (80% errors), in which the first digit was smaller than 5 (e.g., 30901, 20302)—indicating that she did not comprehend the decimal structure of these longer numbers, and therefore estimated the size of the number only on the basis of the identity of the leftmost digit (e.g., when she read 30901, she may have created a 4-digit decimal frame for it, starting with three-thousand, and hence judged it as smaller than 5500).

Nomi had no errors with the 3- and 4-digit numbers. She made significantly more errors on this test than the controls (who made only 3% errors, SD = 7%, Crawford and Howell's $t(9) = 2.40$, $p = 0.02$).

4.6.2. Intact Comprehension of Signed Numbers: >5500

We used the same task as in 4.6.1, but this time the numbers were signed to Nomi rather than presented in writing.

Method

The task included the same numbers as in the previous section (4.6.1), and additional 10 numbers of different lengths—a total of 30 numbers (13 of them were 5-digit numbers, and in eight of these the leftmost digit was smaller than 5). Nomi was presented with a signed multidigit number and was requested to decide, for each number, whether it was greater or smaller than 5,500.

Results

Nomi made only one error in this task, on the first number presented in the task (5601). She responded correctly to all other numbers, including for the eight 5-digit numbers in which the first digit was smaller than 5—the type of numbers that she missed in the written version of the test.

These two >5500 tasks together indicate that Nomi has a similar deficit in comprehension to the one she has in reading-then-signing when the numbers are written, but she does not have a deficit in understanding these numbers when they are signed rather than written to her. This suggests that it is not the comprehension processes themselves that are impaired but rather getting to these processes from written Arabic number input.

4.6.3. Impaired Performance in a Number Comparison Task

We further assessed Nomi's comprehension of numbers using a number comparison task, in which she was requested to decide which of two multidigit numbers is greater.

Method

The task included 68 number pairs, in 40 of which the numbers differed in length, and in 28 the numbers were in the same length. Of the different-length pairs, 32 were such that the first digit of the longer number was *smaller* than the first digit of the shorter number (e.g., 6493 and 52879, henceforth: the "incompatible condition"). The incompatible condition included 16 pairs of a 4- and a 5-digit number, and 16 pairs of a 5- and a 6-digit number. Eight other pairs were of the 'compatible' condition, in which the first digit of the longer number was also *greater* than the first digit of the shorter number.

Of the same-length pairs, 20 (8 pairs of 4-digit numbers, 8 of 5-digit numbers, and 4 of 6-digit numbers) differed already in the first, leftmost, digit, and 8 (3 pairs of 4-digit numbers, 3 of 5-digit numbers, and 2 of 6-digit numbers) differed only starting the second or third digit.

The numbers in each pair were presented to Nomi written one next to the other with 5 spaces between them. She was requested to decide, for each pair, whether the right or the left number was greater by pressing either the right or the left arrow on the keyboard. The pair remained in front of her until she pressed a key, without time limit. The test began with four practice items.

Results

We did not press Nomi for response times in this task, and her RTs were generally longer than those of the control participants, though not significantly (Nomi: M = 3003 ms, SD = 1,490; control M = 2188 ms, SD = 703; RT for error responses and outliers of >3SD from the participant's mean and RTs for errors were excluded). This probably resulted in her not making many errors (she made 3 errors, all of them in the longer numbers: one in a pair of 5- and 6-digit numbers of the "incompatible" condition, and two in pairs of 6-digit numbers with a different first-digit. The number of errors she made was significantly larger than that of the controls (M = 1.30, SD = 0.82, Crawford and Howell's $t(9) = 1.98$, $p = 0.04$), but still relatively low.

Very interesting findings, however, emerged from her pattern of response times in the various conditions. The controls had similar RTs for both types of different-length pairs (the compatible and incompatible conditions), and in fact, 4 of the 10 control participants even had lower average RTs for the incompatible condition, resulting in a relatively small difference between them[2] (Mean difference = 99 ms, SD = 277). In marked contrast, Nomi had much longer RTs on the incompatible condition, and the difference between her average RTs in the incompatible and compatible conditions was significantly larger than that of the controls (Nomi's mean difference = 643 ms, Crawford and Howell's $t(9) = 1.87$, $p = 0.047$, **Figure 6**).

This suggests that whereas typical readers can perform number comparisons on the basis of decimal structure, and therefore are less affected by the first digit when there is a difference in decimal structure, Nomi found it more difficult to rely on the decimal structure of the longer numbers (of 5- and 6-digit numbers) and was therefore more affected by the first-digit in these numbers (possibly because she was relying on the identity of the digits, and specifically on that of the leftmost ones). This type of stimuli was exactly the type that was difficult for her in

[2]This compatibility effect may be in line with other compatibility effects of irrelevant digits (e.g., Nuerk and Willmes, 2005), and with other effects of irrelevant properties on RTs (e.g., Domahs et al., 2010, 2012) in size judgment tasks.

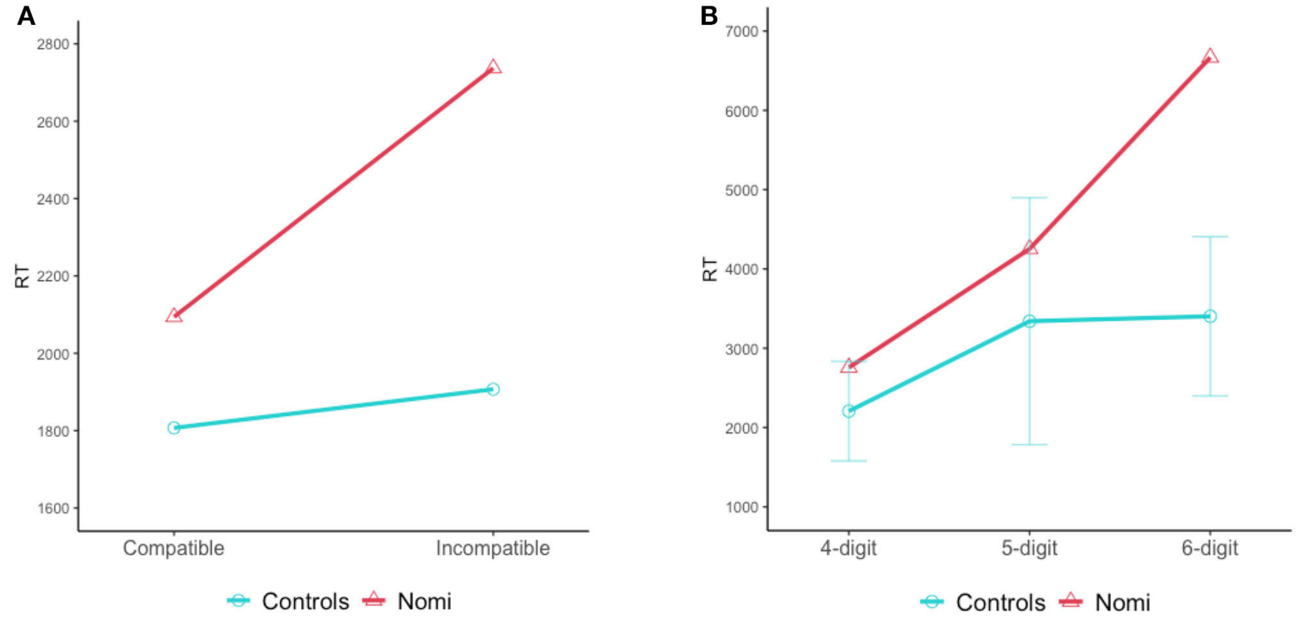


FIGURE 6 | Number-comparison task: Mean RTs of Nomi and the control group. **(A)** RTs in the different-length pairs in the compatible condition (in which the longer number also had a larger first-digit) and in the incompatible condition (in which the longer number had a smaller first-digit). **(B)** RTs in the same-length pairs with similar first-digit, of 4-, 5-, and 6-digit numbers.

the >5500 task (Section 4.6.1) as well (numbers longer than 5500 but with a first-digit smaller than 5).

In addition, Nomi's RTs were significantly longer than the controls' in the same-length conditions in pairs of 6-digit numbers, which was most prominent with 6-digit numbers that had the same first-digit and differed only in the second- or third-digit (controls: M = 3,403 ms, SD = 1,004 ms; Nomi: M = 6,667 ms, SD = 726, Crawford and Howell's $t(9) = 3.10$, $p = 0.006$). As depicted in **Figure 6**, her RTs in the same-length 4- and 5- digit numbers (same first-digit) were similar to the controls' and far shorter than her RTs for the 6-digit numbers.

This forms another indication of her difficulty in processing of long numbers of 6-digits, when building the decimal frame is necessary, and yet another indication that impairment of decimal-frame-construction for longer numbers affected not only her reading, but also her comprehension.

4.7. The Deficit Is in Creating a Number Frame From Written Numbers: Intact Number Production and Comprehension When the Decimal Positions Are Provided

4.7.1. Reading the Same Multidigit Numbers Presented With a Comma Separator

If Nomi's decimal shifts and decimal structure errors in reading-then-signing originate in a deficit in parsing the number into a number frame, adding a comma separator should help her parse the number into triplets, which would help her in creating the appropriate number frame, so she is expected to make fewer decimal shifts than in reading numbers without a comma (see Dotan and Friedmann, 2019, for similar rationale).

Method

Nomi read aloud the same 60 numbers that she had read in the baseline task, but here the numbers were presented with a comma separator between the thousands and hundreds digits (e.g., *12,592*, whereas in the experiment in the baseline task it was *12592*) (multidigit number reading with comma separator A, MAYIM battery, Dotan and Friedmann, 2014). Just like in the baseline task, the numbers were written one above the other, and Nomi was asked to read each number and then sign it in ISL.

Results

Nomi made only 2 decimal errors in reading the numbers with a comma separator—one with a 4-digit number, and the other with a 6-digit number—and she immediately self-corrected both of these errors. In contrast to her performance when reading these numbers without a comma separator (in the baseline task), here she was able to read all numbers without giving up and declaring she could not sign them, even in the longest numbers. Her reading of the same 5–6 digit numbers with a comma separator was significantly better than her reading of these numbers without a comma (McNemar test $p = 0.0001$). In addition, Nomi reported that reading the long numbers with a comma separator was much easier for her than reading them without it.

This result supports the conclusion that her deficit was related to the conversion of the number and parsing it into the decimal frame, and it also lends further support for the conclusion that she had no phonological output deficit.

4.7.2. Intact Comprehension of Written Numbers With a Comma Separator: > 5500

Method

Nomi was presented with the same multidigit numbers as in the test in the written >5500 test (Section 4.6.2). The numbers were printed scattered on a sheet, however, this time they appeared with a comma separator between the thousands and the hundreds digits. Nomi was asked to mark all numbers that were greater than 5500.

Results

Nomi marked correctly all the numbers >5500, including the 5-digit numbers with the first-digit smaller than 5, which she missed in the version written without the comma separator.

These two tests indicate that when an indication as to the decimal structure is given in the written number, reading and comprehension improve considerably and the decimal errors almost disappear. Another result we reported above in section 4.5.1 supports the same point: when Nomi read multidigit numbers presented as Hebrew number-words rather than as written Arabic numerals, she signed them without decimal errors. Notice that the Hebrew number-words provide the abstract decimal structure of the number but not the decimal word-frame of the verbal-signed number in ISL. Nevertheless, once the decimal structure was provided to her (in a separate system- a written spoken language), she did not make decimal errors. A summary of Nomi's performance in all multidigit number tasks is given in **Table 1**.

4.8. Dissociation Between Number Reading and Word Reading

We have established that Nomi has a deficit in the conversion of the written number to its decimal frame. It is interesting to examine whether her deficit is selective to number-reading or whether she also has a deficit in reading words and converting them to their verbal representation.

Dotan and Friedmann (2019) reported double dissociations between dysnumeria in various loci in the number-reading model and dyslexia in reading words in parallel word-reading components. Here we examine whether such a dissociation can also be found between reading-then-signing written numbers and reading-then-signing written words.

4.8.1. Reading and Signing Written Hebrew Words

Method

Nomi was asked to read-then-sign a total of 406 written Hebrew words, presented in 5 tests (adapted from the TILTAN battery, Friedmann and Gvion, 2003). All these tests included lists of single words, presented one above the other. Nomi was requested to read each word and sign it in ISL. Below we describe each of the five tests. Nomi's reading performance was compared to control groups of hearing adults (see the number of participants in each control group and their average performance on each test in **Table 2**); In two of the reading tests, her performance was also

TABLE 2 | Nomi's performance (percentage correct performance and error types) and the control groups' performance in the Hebrew word reading tasks.

Test	Sensitive to…	Total words	% correct performance	Error types	Deaf signer controls M(SD)	Hearing controls M(SD)
Read-then-sign tasks						
Screening: single words	Various types of dyslexia	92	98%	1 morphological error, 1 low-frequency Hebrew word that she did not know how to translate into sign	–	*N* = 1,045 98.1% (2.0%)
Migratable words	Letter position dyslexia	160	96%	3 semantic errors (signing a sign semantically related to the target word), 2 letter migrations, 1 omission of a double letter, 1 morphological error	*N* = 7 96% (2%)	*N* = 698 96.6% (2.3%)
Homophone and potentiophone	Surface dyslexia	86	99%	1 letter omission	*N* = 6 97% (3%)	*N* = 257 93.7% (4.1%)
Long words (5–11 letters)	Buffer deficits	36	100%	–	–	*N* = 35 97.6% (2.4%)
Morphologically complex words	Morphological difficulties	32	100%	–	–	*N* = 20 97.5% (1.7%)
Same-different task						
Same-different task	Letter position dyslexia, letter identity dyslexia, and word length	40 different pairs	100%	Did not miss any difference between words	–	*N* = 24 97%, 3%

compared to a group of 7 deaf native ISL signers (4 women and 3 men, aged 18-44). The five tests were:

1) **Tiltan Siman screening**: 92 words sensitive to different types of dyslexia (including migratable words, irregular words, long words, morphologically-complex words, function words, abstract words, words with many orthographic neighbors, words with double letters and more).
2) **Migratable word list**: 160 migratable words, in which a transposition of two letters creates another existing word (e.g., *stakes,* which can be read with a transposition as "skates," Friedmann and Gvion, 2001; Friedmann et al., 2010).
3) **Surface list**: 86 potentiophonic and homophonic Hebrew words. Potentiophones are words that if read via the sub-lexical route can potentially be read as other existing Hebrew words, which sound differently (e.g., *none*, which may be read via grapheme-to-phoneme conversion as "known," Friedmann and Lukov, 2008; homophones are words that sound the same but have different meaning and spelling, so they translate to different signs).
4) "**Artichoke"- long and morphologically complex words**: 36 long and morphologically complex Hebrew words of 5–11 letters.
5) "**Duvshaniyot**"- **morphologically complex words** with derivational morphology: 32 short and long words with Hebrew derivational morphology, in which the same Hebrew root put in a different morphological pattern leads to a change in meaning (e.g., *paired* and *impair*). This type of words should elicit errors in translation to signs if the morphological template is incorrectly identified. If the structural process that creates a multidigit number frame is similar to the process creating the morphological structure of a word, then a dissociation in this task would be especially telling.

Results

Nomi's word-reading was excellent, even compared with hearing Hebrew readers, and very different from her number-reading. Nomi's performance in the five word-reading tasks as well as the analysis of the (few) errors she made in each of the tests are summarized in **Table 2**. In total, in reading 406 words with various kinds of complexity, Nomi made only 10 errors (98% correct), a rate that is well within the range of typically-hearing control adults. For each of the tasks, her performance was similar to that of hearing controls, and to that of the deaf signers, in both groups her performance puts her in the higher range of performance ($p > 0.05$ for all comparisons between her performance and the control groups using Crawford and Howell, 1998, *t*-test).

Three types of errors in word-reading may be parallel to the errors Nomi made in reading numbers: morphological errors, letter transpositions, and length errors, manifesting in letter omissions or additions. As summarized in **Table 2**, she did not make any of these errors in a rate higher than the controls:

1) she made only 2 morphological errors in all 406 words she read, and no morphological errors in the Artichoke and the Duvshaniyot tests, which were created to examine the reading of morphologically complex words. This indicates that she does not have a general problem with reading structurally complex items, but rather a deficit that is limited to the structure of (long) multidigit numbers.
2) In the migratable words test, which was created to elicit letter migrations within words, as all the 160 words in the test were "migratable," she made only 2 letter migrations, again, better than the average performance of hearing readers.
3) In the long-word test she made no reading errors at all, and she did not make more letter omissions or additions than the typically hearing controls, indicating that her length perception of words was unimpaired too.

   So not only the general percentage correct performance but also the analysis of her error types indicates a clear difference between her word reading and her number-reading.

Additionally, in the Artichoke and the Duvshaniyot tests, Nomi made no errors at all, not even in the longest words. Comparing her reading of 5–6 letter words to 5–6-digit numbers (a total of 5 errors out of 190 words, compared to 26 errors out of 54 numbers), her word reading was significantly better, $\chi^2(1) = 74.50$, $p << 0.0001$) when taking into account all words of 5 letters or longer yielding a total of 6 errors out of 217 5–11 letter words, this difference is even greater, $\chi^2(1) = 81.22$, $p << 0.0001$). This indicates a clear dissociation between her reading of long multidigit numbers and her reading of long words, and shows that these two processes are separate.

4.8.2. Same-Different Decision: Pairs of Written Words Presented Together

Method

To assess Nomi's orthographic-visual analyzer and to further compare her word- and number-reading, we administered a same-different task with pairs of written words. Nomi was presented with 60 pairs of Hebrew words printed one next to the other in a list, and was asked to mark the pairs in which the two words were identical. The test included 10 words that differed in length (length difference created by doubling a letter in one word that created the other word, similar to driver-diver or diner-dinner), 10 that differed in one letter identity, 20 that differed in the order of two adjacent letters (e.g., flies-files, skates-stakes), and 20 identical pairs.

Results

Nomi did not miss any difference between words (she did not mark any pairs that were not identical), and she only missed 2 pairs of identical words (3% of the total number of words, which falls well within the results of hearing controls ($n = 24$, mean errors = 3%, SD = 3%, Lorber, 2020). This supports that she has very good reading, and she does not make errors of letter-identity, migrations of letters within the words, or omissions of double letters which would indicate number length deficit.

4.8.3. Interim Summary: Words vs. Numbers

To summarize, Nomi's word-reading was very good. It was similar to that of the higher range of hearing controls and of deaf native signer controls. Her general performance in the word-reading tasks, then, points to a clear dissociation between her good word-reading and her poor number-reading. We also found that the component that was impaired in her number-reading did not cause a parallel impairment in word-reading: if processes related to the decimal structure of numbers are parallel to word structure –morphology, she did not make morphological errors in reading words, indicating that the structural component is not shared by numbers and words.

Deficits in the conversion of number-length could be envisaged also as parallel to deficits in word-length, which should have led to errors of letter addition or omission, which may be similar to number-length errors, and errors of letter-position which may be similar to decimal shifts in reading numbers. Here, too, Nomi did not make such errors more than the controls.

Therefore, in each of these subprocesses she showed intact performance in word-reading: she made no morphological errors in the two tests with morphologically complex words, she made only two letter omissions of the 406 words she read, she made fewer letter transpositions than the hearing control average in the migratable word test, and she did not miss any difference in letter-order or in word-length in the same-different task, indicating that there was no shared component that was impaired, and that her impairment was selective to number-reading.

5. DISCUSSION

We brought here the first report of a specific type of dysnumeria—an impairment in number-processing, in a deaf signer, Nomi. Nomi's dysnumeria manifested itself in a difficulty to read and comprehend the decimal structure of long multidigit Arabic numbers.

Her most prominent error type in multidigit number reading was decimal errors—she had difficulty in processing the decimal structure of the number, most notably in longer numbers of 5–6 digits.

5.1. Nomi's Functional Locus of Deficit in the Number Reading Model

We suggest that Nomi's impairment lies in the conversion of written multidigit numbers into the abstract decimal frame, namely in the connection between the (intact) numeric-visual analysis stage of extraction of decimal structure and the decimal structure component. This deficit is marked (2) in **Figure 7**, which already uses a modified model that will be motivated and explained below. As we discuss below, we assume that the construction of the non-verbal decimal frame of the number serves both reading aloud and comprehension of written numbers, and therefore a deficit in the connection from visual input to this process affects both reading aloud and comprehension from written input.

Nomi showed difficulties in reading multidigit numbers, mainly in longer numbers of 5–6 digits. Her errors were mostly decimal shifts of the leftmost digit (e.g., reading Arabic numbers such as 34567 as three thousand…), resulting in saying the first digits in an incorrect (and smaller) decimal position.

As depicted in **Figure 7**, decimal shift errors could, theoretically, result from three general stages in the number-reading process: the numeric-visual input processes of extracting number-length ①, the conversion of a written number to its abstract decimal frame ②, or the verbal-output processes ③–⑥: in linearization of the abstract frame into a verbal number frame ③, in misalignment of the ordered digits into the number frame during the binding process ④, in the retrieval of number-words/signs (retrieving/producing a word with incorrect class) ⑤, or in holding too many number-words/signs at the same time ⑥.

We will now use the results of the tests reported above to show how we reached the conclusion that Nomi's deficit is in the conversion of written numbers to the decimal frame and why the other theoretically possible loci are excluded for her deficit. **Figure 7** summarizes all the test results upon which we base our conclusions about spared and impaired components, which we describe and discuss in detail below (blue for good performance, orange and crimson for impaired performance).

The early numeric-visual analysis stage is ruled out as the source of Nomi's deficit on the basis of her good performance in tasks that involve the numeric-visual analysis without the later conversion and phonological output stages. She performed very well in the three same-different tasks, and specifically in detecting pairs that differed in number-length or in digit-order, suggesting she did not have deficits in extracting number-length and digit-order from written numbers. She also performed well in digit-sequence decision. Her good performance in reading the same numbers when signing them non-decimally, digit-by-digit or triplet-by-triplet, provides further strong evidence that her numeric-visual analysis was intact. Her deficit emerged only when she had to use this information to read the number "aloud" as a multi-digit number, namely, when she had to create the decimal number frame for the written number (for production, and, as we will show below, also for comprehension). A deficit in the numerical input buffer is also excluded in view of the absence of order- and identity-errors in all input tasks.

The verbal output processes are also ruled out as the source of Nomi's decimal errors: when she produced long multidigit numbers in tasks that did not involve written Arabic numerals (in reading numbers written as Hebrew number words and in repeating multidigit numbers), she made no decimal errors. She could sign without any decimal errors long multidigit numbers written in Hebrew words (of the same length as she failed to sign in reading). Multidigit number repetition was not easy for her, but still, she made almost no decimal errors in two tasks of multidigit number repetition. These results demonstrate that Nomi can produce multidigit numbers with their correct decimal structure when no Arabic numeral reading is involved.

Several additional findings support the conclusion that her production is intact. When she read the same multidigit numbers with a comma separator, she made significantly fewer decimal errors. Reading numbers with comma requires the same output processes as reading numbers without comma, so her difficulty in numbers without comma could not have resulted from a deficit in the output processes.

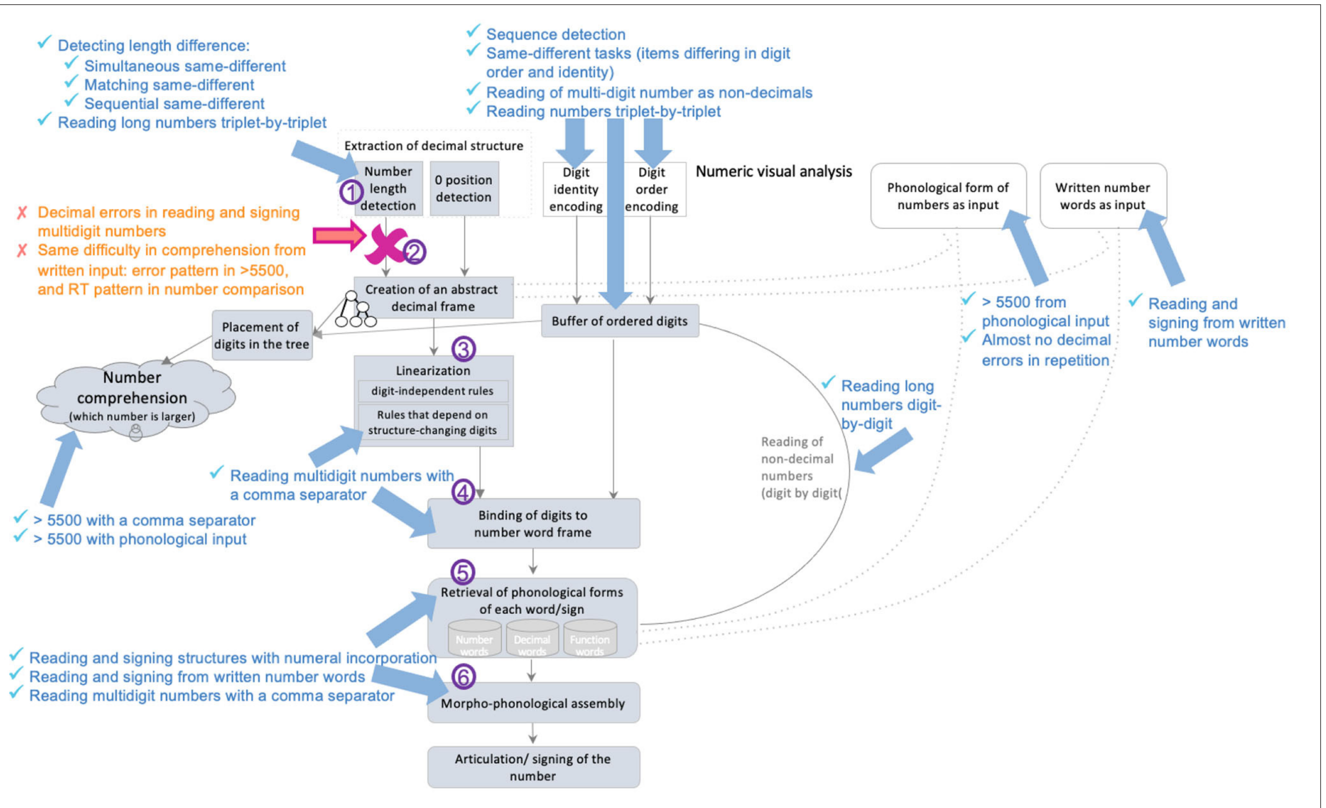


FIGURE 7 | The number reading model with Nomi's performance related to each of the components. In blue: performance that is within the control range; orange: impaired performance, lower than the control group. Circled numbers indicate theoretically possible sources of decimal errors in multidigit number reading. Crimson X indicates the locus of Nomi's deficit in view of her performance in the various tasks.

Another indication of her intact phonological output buffer is her good reading-then-signing of structures involving number incorporation (like "twice"), as number incorporation takes place in the phonological output buffer (Haluts and Friedmann, 2020).

Her good production of number incorporation, which are morphologically-complex structures involving numbers, also demonstrates that her decimal errors did not stem from a deficit in the processing of morphologically complex structures that involve numbers. This conclusion is also supported by her good production of the morphologically complex decimal numbers in repetition and reading-then-signing of Hebrew number words.

Finally, in reading multidigit numbers with fewer number-words (e.g., 400300), she still made many decimal errors, in a rate similar to the error rate she had in reading multidigit numbers with many different digits. These numbers include the same number of written digits but require far fewer number-signs in production, so these numbers should have made a difference only for the phonological output stages, and should have been easier had the deficit been in the phonological output stages. The finding that Nomi still made many decimal errors in reading these numbers suggests that the source of her decimal errors was not the need to hold and assemble many number-signs together.

The pattern of Nomi's errors can be taken as additional evidence that Nomi's decimal errors did not result from a phonological output deficit: had her deficit been in the retrieval of number-signs in the correct class, we would not expect errors only on 5–6-digit numbers, and we would not expect errors mainly in the leftmost digit, and toward a smaller decimal position. Additionally, had she had a deficit in lexical retrieval, we would expect other phonological substitutions, such as other number-identity errors (signing THREE instead of FIVE), which almost never occurred.

We therefore conclude that Nomi's deficit in number-reading (reading-then-signing) lies in the creation of the decimal number frame from written number input. That is, in the conversion of the decimal information extracted in the (intact) numeric-visual analyzer into a number frame. The findings that: (a) Nomi made errors only in numbers of 5-digits and up, and had almost no decimal errors in 4-digit numbers and no errors in 3-digit numbers, and (b) her decimal errors were always in the direction of using a number frame that is smaller than the written number (in one decimal position) suggest that her deficit was a result of a limitation in the size of the number frame into which she could place the written numbers. She is able to build a smaller tree/number frame from written input, so the frame for shorter numbers (of 4 digit or less) is created correctly. However, for 5- and 6-digit numbers she cannot create a frame that would be suitable for the written input. She typically created a 4-digit frame for the 5-digit numbers, and never succeeded to create decimal structures for 6-digit numbers, mainly saying "I can't," starting and failing to create a 5-digit tree, or breaking the number into two triplets. This treelet deficit might be similar to Power and Dal Martello (1997) study that showed, for 7-year-olds, decimal errors that result from their inability to create larger trees. Nomi's impairment pattern is possibly similar also to that of ED (Dotan and Friedmann, 2018, 2019), who also had decimal errors, and mainly on the longer numbers, but ED made 26% errors in reading 4-digit numbers, so she might have had difficulty with even smaller decimal frames compared to Nomi.

This deficit can be conceptualized either as a deficit in the conversion from the numeric-visual analysis stage to the number frame, or as a deficit in a component of number frame building for written numbers[3]. At this point we do not see a way to distinguish between decimal-frame construction that is input-specific (written numbers, phonologically presented numbers), and a single decimal-frame construction component that has separate connections from the different inputs.

5.2. Nomi's Multidigit Number Comprehension and Its Implication for the Model

Multidigit number comprehension, as measured by number-comparison tasks, was impaired along the same lines as Nomi's reading "aloud": just like her decimal errors in reading 5–6-digit numbers, she also had difficulty in comprehending the decimal structure of numbers of this length. Whereas her comprehension of short, 3–4-digit numbers was intact, she failed to detect that 5-digit numbers were larger than 4-digit numbers when their first digit was not larger than the first digit of the 4-digit numbers (23675>5500) in two different tasks. This difficulty manifested itself both in errors (in the >5500 task) and in response time patterns. This suggests that her deficit in processing the decimal structure of the written numbers also affected her comprehension of number-size.

What is the source of this deficit in comprehension? The data showed that her early numeric-visual stage of number-length encoding (as well as digit-identity and order) was intact. So just as the deficit in reading "aloud" could not have emerged from the numeric-visual stage, neither could the comprehension deficit. Her comprehension of the same numbers from signing was also intact, indicating that the deficit was not in the comprehension itself but was limited to written numbers.

[3]It is also possible to conceptualize the locus of this deficit as a deficit in number length analysis of the numeric-visual analysis system in a process that **only** applies to reading aloud and comprehension of multidigit numbers with their decimal structure (but not to same-different decision tasks). We find such conceptualization inferior to a deficit in the connection between the analyzer and the decimal frame building, because it requires assuming that the number length analysis stage functions differently in different tasks and "knows" already what the task is and whether it requires multidigit numbers as phonological output (or size comprehension) or not. It also requires that the early length detection stage in the numeric-visual analysis includes a function that serves late stages such as reading aloud and comprehension but not same-different decision. Assuming that the deficit resides in the connection of the length detection component to the frame building component does not require such less-probable assumptions.

It therefore seems that the source of the deficit in number comprehension was impaired conversion of the written number into its decimal frame, just like in number-reading. This suggests the possibility that the deficits in reading and in comprehension are in fact a single deficit, at a component that is shared by the two processes. Conversely, it might be that there are two independent identical impairments of converting the numeric-visual information of multidigit numbers into their decimal form—one in conversion for reading, the other in conversion for comprehension, and both manifest themselves in numbers of the same lengths and in similar errors. Whereas this two-deficit option is possible, Ockham's razor drives us to prefer the option of a deficit in a component shared by reading and comprehension. We therefore suggest that Nomi's deficit lies in the conversion of information from the numeric-visual stage into the shared decimal number frame construction stage, thus affecting both reading "aloud" and comprehension.

5.3. Implications for the Number Reading Model

Nomi's pattern of impairment thus offers new insights to the number-reading model. The first insight relates to the number-frame-building component shared by reading aloud and comprehension, described in the previous section. In **Figure 8** we provide a possibility for the architecture of such shared component: the creation of the number frame is abstract and non-verbal and is shared by reading and comprehension. This abstract, non-verbal, number frame creation component provides the information of the decimal structure of the number both for later verbal stages (in signs or spoken-words) and for comprehension (in our case, number comparison), so it is connected to the linearization component, which is verbal and language-dependent, and to the further comprehension components.

The idea of a ***shared abstract decimal-structure component*** for comprehension and production is in line with McCloskey's (1985, 1986, 1992) idea that reading aloud of multidigit numbers passes through an abstract stage that is shared with comprehension. Differently from McCloskey, and in line with Cohen and Dehaene (1991) and Dehaene and Cohen (1995), we do not assume that this shared component is a semantic representation that follows, and requires, "number comprehension," but rather a stage that immediately follows numeric-visual analysis, preceding both production and comprehension of written multidigit numbers. We suggest that this stage involves the construction of an abstract, non-verbal, decimal number frame.

Once we assume such an abstract number frame component, we can assume it is responsible both for the decimal structure and for the parsing into triplets. We currently do not see the need to assume a separate triplet-parsing component at the numeric-visual analysis stage, and patients like ED (Dotan and Friedmann, 2018, 2019), who showed impaired triplet-parsing, may be impaired in the general process converting the information from the numeric-visual analysis onto the number frame (at the moment we remain agnostic as

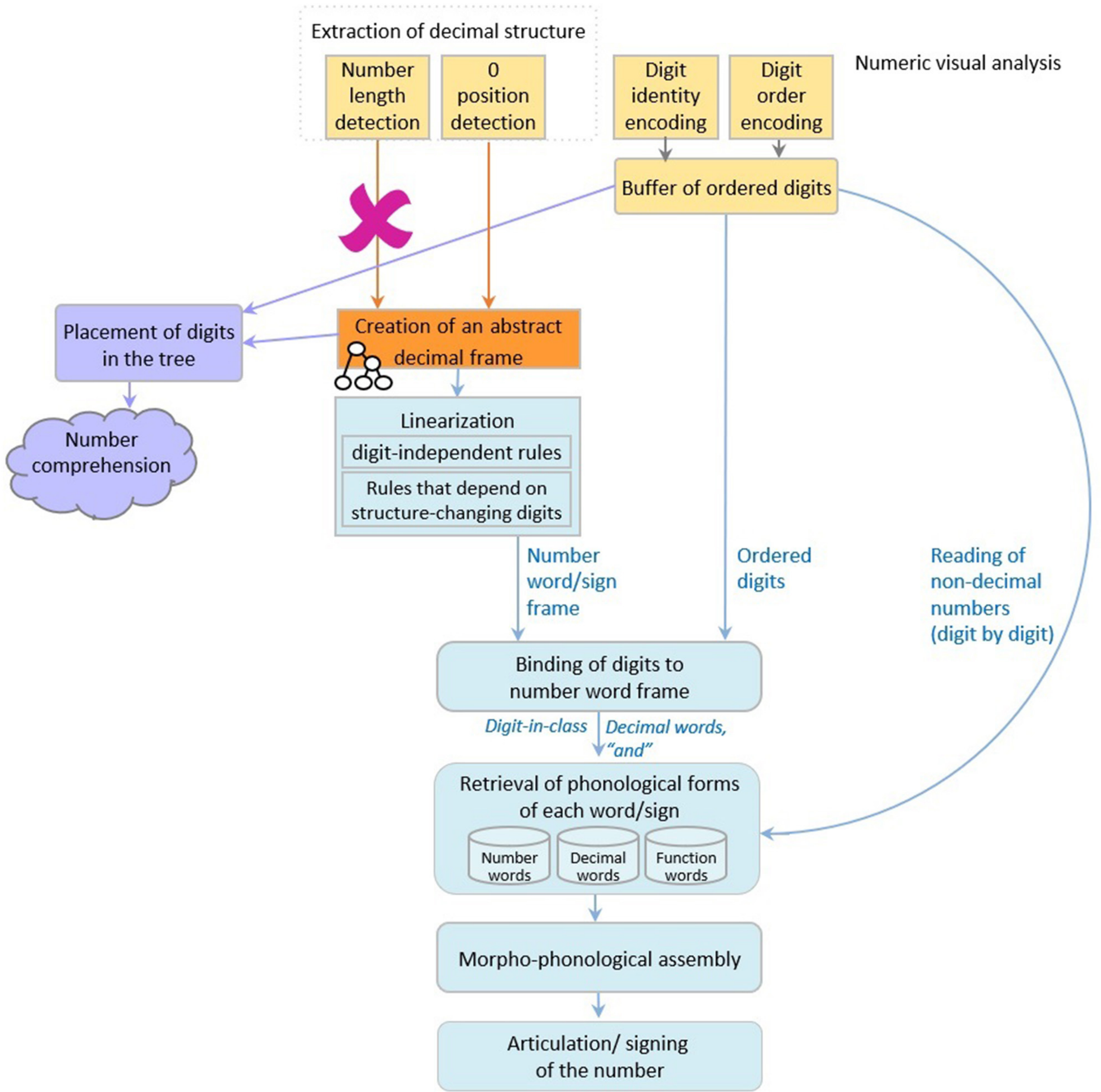


FIGURE 8 | A revised number reading model. Three modifications to Dotan and Friedmann's (2018) model: (1) The decimal structure building component is non-verbal, and it is shared with comprehension processes, (2) an addition of a non-decimal number reading route, (3) triplet parsing is no longer part of the numeric-visual analysis stage.

to whether there is a separate triplet-parsing component, whether it resides in the numeric-visual input or in the phonological output components or both, and whether it depends on the way the target language divides numbers into groups).

A second insight from this study is the existence of ***a non-decimal route***, which allows a digit-by-digit reading of multidigit numbers, without forming a decimal representation (somewhat similar to the non-lexical route in word-reading, Coltheart et al., 2001). Nomi had a deficit in reading multidigit numbers with decimal structure, but had no problem reading the same numbers digit-by-digit. A similar pattern was reported for the patients in Cohen and Dehaene (1995), but for digit-identity errors rather than decimal errors. We suggest that this digit-by-digit reading is performed in the non-decimal route, portrayed in **Figure 8** with an arc connecting the numeric buffer of ordered digits with the phonological output component, which bypasses the decimal structure construction.

5.4. Dissociation With Reading Words

In marked contrast to her impaired multidigit number reading, Nomi's word reading was intact. Her performance in reading 406 words did not differ from that of typically-hearing Hebrew speaking adults, nor did it differ from the reading of deaf signers. This already shows a clear dissociation between her poor multidigit number reading and her very good word-reading.

Additionally, when one examines the pattern of errors that would have been expected had she had a deficit in word reading that is parallel to her deficit in number-reading, it is clear that she does not make similar error types: she made only two letter omissions, two letter transpositions, and two morphological errors (1.5% together) which could be counted as parallel to decimal errors in numbers, compared to 20 out of 100 (20%) decimal errors in her number-reading.

Additionally, her deficit in number-reading was most pronounced in the longer, 5–6-digit numbers. Conversely, her word reading of 5- and 6-letter words (and even longer words) was unimpaired and significantly better than her reading of numbers of similar lengths.

These results, thus, support the conclusion that number-reading is implemented, at least in part, by mechanisms that are different and separate from the ones that are used in word-reading (Friedmann et al., 2010; Shum et al., 2013; Abboud et al., 2015; Hannagan et al., 2015; Güven and Friedmann, 2019; for a review, see Dotan and Friedmann, 2019).

5.5. A Specific Impairment in Number Reading in a Deaf Signer

Nomi, a deaf user of a sign language, showed a specific impairment in number-reading. As discussed above, Nomi's impairment resulted from a very specific stage in the number-reading model. Her impairment seems similar to that of NR, reported by Noel and Seron (1993) and discussed above in the introduction, and to that of ED, reported by Dotan and Friedmann (2018, 2019). The fact that very similar number-reading impairments can be found in spoken language users and in sign language users suggests that the mechanisms that process number reading are shared by all speakers of human languages and do not depend on the modality in which the language is transmitted.

5.6. Conclusion

We reported here of the first in-depth investigation of a selective dysnumeria in a user of a sign language. Her pattern of errors and performance in various tasks indicated a decimal-structure conversion dysnumeria, a deficit in the construction of decimal number frame from written numbers of 5-digits or longer. Her deficit was shared by reading aloud (reading-then-signing) and comprehension processes. Nomi made no errors in tasks that did not require the construction of decimal frames from written numbers: she performed well in tasks involving only the numeric-visual analysis, and made virtually no decimal errors in tasks involving the production of multidigit numbers without written input. When she read the same long multidigit numbers with cues as to the decimal structure, she made fewer decimal errors, and when she read the exact same numbers in a non-decimal system in ISL, which only involves digit-by-digit signing, she made no such errors. These results indicate that prior to the construction of the verbal number frame, a non-verbal abstract frame is constructed, which is shared by reading and comprehension. Additionally, these results provide evidence for a parallel, non-decimal reading route for reading multidigit numbers. The assessment of dysnumeria in sign language, thus, opened a new window to insights regarding the number-reading process.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available due to privacy issues. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tel Aviv University Ethics committee. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

NF, NH, and DL worked together on the conceptualization, creation of tests, testing Nomi, and analyzing the results. NF and NH worked together on all stages of writing. NF provided resources and funding. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Human Frontier Science Program: RGP0057/2016; and by the European Union's Horizon 2020 Research and Innovation Programme SIGN-HUB: H2020-REFLECTIVE-2014-2015 grant 774 agreement No. 693349.

ACKNOWLEDGMENTS

We thank Nomi for her enthusiastic and patient participation even in very boring tasks, we thank Maor Ben Zeev for discussions of various ISL structures and for his precious help in the collection of control data, and we thank Dror Dotan for discussions of the model.

REFERENCES

Abboud, S., Maidenbaum, S., Dehaene, S., and Amedi, A. (2015). A number-form area in the blind. *Nat. Commun.* 6:6026. doi: 10.1038/ncomms7026

Austin, G. F. (1975). Knowledge of selected concepts obtained by an adolescent deaf population. *Am. Ann. Deaf.* 120, 360–370.

Basso, A., and Beschin, N. (2000). Number transcoding and number word spelling in a left-brain-damaged non-aphasic acalculic patient. *Neurocase* 6, 129–139. doi: 10.1080/13554790008402766

Biran, M., and Friedmann, N. (2004). *SHEMESH: Naming a Hundred Objects*. Tel Aviv: Tel Aviv University.

Blanken, G., Dorn, M., and Sinn, H. (1997). Inversion errors in Arabic number reading: Is there a nonsemantic route? *Brain Cogn.* 34, 404–423. doi: 10.1006/brcg.1997.0917

Bull, R., Marschark, M., Sapere, P., Davidson, W. A., Murphy, D., and Nordmann, E. (2011). Numerical estimation in deaf and hearing adults. *Learn. Individ. Differ.* 21, 453–457. doi: 10.1016/j.lindif.2011.02.001

Cappelletti, M., Kopelman, M. D., Morton, J., and Butterworth, B. (2005). Dissociations in numerical abilities revealed by progressive cognitive decline in a patient with semantic dementia. *Cogn. Neuropsychol.* 22, 771–793. doi: 10.1080/02643290442000293

Cipolotti, L., and Butterworth, B. (1995). Toward a multiroute model of number processing: impaired number transcoding with preserved calculation skills. *J. Exp. Psychol.* 124, 375–390. doi: 10.1037/0096-3445.124.4.375

Cipolotti, L., Warrington, E. K., and Butterworth, B. (1995). Selective impairment in manipulating arabic numerals. *Cortex* 31, 73–86. doi: 10.1016/S0010-9452(13)80106-2

Cohen, L., and Dehaene, S. (1991). Neglect dyslexia for numbers? A case report. *Cogn. Neuropsychol.* 8, 39–58. doi: 10.1080/02643299108253366

Cohen, L., and Dehaene, S. (1995). Number processing in pure alexia: the effect of hemispheric asymmetries and task demands. *Neurocase* 1, 121–137. doi: 10.1080/13554799508402356

Cohen, L., Verstichel, P., and Dehaene, S. (1997). Neologistic jargon sparing numbers: a category-specific phonological impairment. *Cogn. Neuropsychol.* 14, 1029–1061. doi: 10.1080/026432997381349

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295X.108.1.204

Crawford, J. R., and Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol.* 12, 482–486. doi: 10.1076/clin.12.4.482.7241

Davis, S. M., and Kelly, R. R. (2003). Comparing deaf and hearing college students' mental arithmetic calculations under two interference conditions. *Am. Ann. Deaf.* 148, 213–221. doi: 10.1353/aad.2003.0018

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N

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

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

Delazer, M., and Bartha, L. (2001). Transcoding and calculation in aphasia. *Aphasiology* 15, 649–679. doi: 10.1080/02687040143000104

Deloche, G., and Willmes, K. (2000). Cognitive neuropsychological models of adult calculation and number processing: the role of the surface format of numbers. *Europ. Child Adolesc. Psychiatry* 9(Suppl. 2), II27–40. doi: 10.1007/s007870070007

Domahs, F., Klein, E., Moeller, K., Nuerk, H.-C., Yoon, B.-C., and Willmes, K. (2012). Multimodal semantic quantity representations: further evidence from Korean Sign Language. *Front. Psychol.* 2:389. doi: 10.3389/fpsyg.2011.00389

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

Dotan, D., and Friedmann, N. (2014). *MAYIM: Battery for Assessment of Numeric Abilities*. Tel Aviv: Tel Aviv University.

Dotan, D., and Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex* 63, 317–351. doi: 10.1016/j.cortex.2014.08.014

Dotan, D., and Friedmann, N. (2018). A cognitive model for multidigit number reading: Inferences from individuals with selective impairments. *Cortex* 101, 249–281. doi: 10.1016/j.cortex.2017.10.025

Dotan, D., and Friedmann, N. (2019). Separate mechanisms for number reading and word reading: evidence from selective impairments. *Cortex* 114, 176–192. doi: 10.1016/j.cortex.2018.05.010

Dotan, D., Friedmann, N., and Dehaene, S. (2014). Breaking down number syntax: Spared comprehension of multi-digit numbers in a patient with impaired digit-to-word conversion. *Cortex* 59, 62–73. doi: 10.1016/j.cortex.2014.07.005

Fischer, S. D., Hung, Y., and Liu, S.-K. (2011). "Numeral incorporation in Taiwan Sign Language." in *Language and Cognition: Festschrift in Honor of James H.-Y. Tai on his 70th Birthday*, ed J. H. Chang (Taipei: The Crane Press), 147–169.

Friedmann, N. (2017). *MILO: Conceptual Odd-One-Out Picture Test*. Tel Aviv: Tel Aviv University.

Friedmann, N., Dotan, D., and Rahamim, E. (2010). Is the visual analyzer orthographic-specific? Reading words and numbers in letter position dyslexia. *Cortex* 46, 982–1004. doi: 10.1016/j.cortex.2009.08.007

Friedmann, N., and Gvion, A. (2001). Letter position dyslexia. *Cogn. Neuropsychol.* 18, 673–696. doi: 10.1080/02643290143000051

Friedmann, N., and Gvion, A. (2002). *FriGvi: Friedmann Gvion Battery for Assessment of Phonological Working Memory*. Tel Aviv: Tel Aviv University.

Friedmann, N., and Gvion, A. (2003). *TILTAN: Battery for the Diagnosis of Dyslexias in Hebrew*. Tel Aviv: Tel Aviv University.

Friedmann, N., and Lukov, L. (2008). Developmental surface dyslexias. *Cortex* 44, 1146–1160. doi: 10.1016/j.cortex.2007.09.005

Frostad, P. (1996). Mathematical achievement of hearing impaired students in Norway. *Europ. J. Special Needs Educ.* 11, 67–81. doi: 10.1080/0885625960110105

Fuentes, M., and Tolchinsky, L. (2004). The subsystem of numerals in Catalan Sign Language: description and examples from a psycholinguistic study. *Sign Language Stud.* 5, 94–117. doi: 10.1353/sls.2004.0022

Genovese, E., Galizia, R., Gubernale, M., Arslan, E., and Lucangeli, D. (2005). Mathematical vs. reading and writing disabilities in deaf children: a pilot study on the development of numerical knowledge. *Cogn. Learn. Diverse Settings* 18, 33–46. doi: 10.1016/S0735-004X(05)18002-1

Girelli, L., and Delazer, M. (1999). Differential effects of verbal paraphasias on calculation. *Brain Lang.* 69, 361–364.

Gottardis, L., Nunes, T., and Lunt, I. (2011). A synthesis of research on deaf and hearing children's mathematical achievement. *Deaf. Educ. Int.* 13, 131–150. doi: 10.1179/1557069X11Y.0000000006

Güven, S., and Friedmann, N. (2019). Developmental letter position dyslexia in Turkish, a morphologically rich and orthographically transparent language. *Front. Psychol.* 10:2401. doi: 10.3389/fpsyg.2019.02401

Gvion, A., and Friedmann, N. (2008). FriGvi: a battery for the diagnosis of phonological working memory. *Lang. Brain* 7, 161–180.

Haluts, N., and Friedmann, N. (2019). *SIMBA: A Battery for Assessment of Phonological Working Memory in Israeli Sign Language.* Tel Aviv: Tel Aviv University.

Haluts, N., and Friedmann, N. (2020). "Signs for a phonological output buffer impairment: Morphological errors in deaf signers with POB impairment," in *Presented at the 56th annual conference of the Israeli Speech Hearing and Language Association* (Tel Aviv).

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., and Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends Cogn. Sci.* 19, 374–382. doi: 10.1016/j.tics.2015.05.006

Iversen, W., Nuerk, H.-C., and Willmes, K. (2004). Do signers think differently? The processing of number parity in deaf participants. *Cortex* 40, 176–178. doi: 10.1016/S0010-9452(08)70940-7

Iversen, W., Nuerk, H. C., Jäger, L., and Willmes, K. (2006). The influence of an external symbol system on number parity representation, or What's odd about 6? *Psychonomic Bull. Rev.* 13, 730–736. doi: 10.3758/BF03193988

Kelly, R. R., and Gaustad, M. G. (2007). Deaf college students' mathematical skills relative to morphological knowledge, reading level, and language proficiency. *J. Deaf Stud. Deaf Educ.* 12, 25–37, doi: 10.1093/deafed/enl012

Korvorst, M., Nuerk, H. C., and Willmes, K. (2007). The hands have it: number representations in adult deaf signers. *J. Deaf Stud. Deaf Educ.* 12, 362–372. doi: 10.1093/deafed/enm002

Liddell, S. K., and Johnson, R. E. (1989). American sign language: the phonological base. *Sign. Language Stud.* 64, 195–277. doi: 10.1353/sls.1989.0027

Lorber, R. (2020). *The effect of Catha Edulis and Cannabis on Language and Cognition.* [Unpublished doctoral dissertation]. Tel Aviv University, United States.

Mantovan, L., Geraci, C., and Cardinaletti, A. (2019). On the cardinal system in Italian Sign Language (LIS). *J. Linguistics* 55, 795–829. doi: 10.1017/S0022226718000658

Marangolo, P., Nasti, M., and Zorzi, M. (2004). Selective impairment for reading numbers and number words: a single case study. *Neuropsychologia* 42, 997–1006. doi: 10.1016/j.neuropsychologia.2004.01.004

Marangolo, P., Piras, F., and Fias, W. (2005). "I can write seven but I can't say it": A case of domain-specific phonological output deficit for numbers. *Neuropsychologia* 43, 1177–1188. doi: 10.1016/j.neuropsychologia.2004.11.001

Mayer, C. (2017). Written forms of signed languages: a route to literacy for deaf learners? *Am. Ann. Deaf* 161, 552–559. doi: 10.1353/aad.2017.0005

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. *Cognition* 44, 107–157. doi: 10.1016/0010-0277(92)90052-J

McCloskey, M., Caramazza, A., and Basili, A. (1985). Cognitive mechanisms in number processing and calculation: evidence from dyscalculia. *Brain Cogn.* 4, 171–196. doi: 10.1016/0278-2626(85)90069-7

McCloskey, M., Sokol, S. M., Caramazza, A., and Goodman-Schulman, R. A. (1990). "Cognitive representations and processes in number production: evidence from cases of acquired dyscalculia," in *Cognitive Neuropsychology and Neurolinguistics: Advances in Models of Cognitive Function and Impairment*, ed A. Caramazza (Hillsdale, NJ: Lawrence Erlbaum Associates), 1–32.

McCloskey, M., Sokol, S. M., and Goodman, R. A. (1986). Cognitive processes in verbal-number production: inferences from the performance of brain-damaged subjects. *J. Exp. Psychol.* 115, 307–330. doi: 10.1037/0096-3445.115.4.307

Meir, I., and Sandler, W. (2007). *A Language in Space: The Story of Israeli Sign Language*. Mahwah, NJ: Psychology Press. doi: 10.4324/9780203810118

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2013). Transcoding abilities in typical and atypical mathematics achievers: The role of working memory and procedural and lexical competencies. *J. Exp. Child Psychol.* 116, 707–727. doi: 10.1016/j.jecp.2013.07.008

Noel, M.-P., and Seron, X. (1993). Arabic number reading deficit: A single case study or when 236 is read (2306) and judged superior to 1258. *Cogn. Neuropsychol.* 10, 317–339. doi: 10.1080/02643299308253467

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

Nuerk, H. C., and Willmes, K. (2005). On the magnitude representations of two-digit numbers. *Psychol. Sci.* 47, 52–72.

Nunes, T., and Moreno, C. (1998). The signed algorithm and its bugs. *Educ. Stud. Math.* 35, 85–92. doi: 10.1023/A:1003061009907

Palma, S., Gubernale, M., Guarnaccia, M. C., and Genovese, E. (2010). Numerical intelligence, verbal competence and intelligence in preschool children with cochlear implants: our findings in a clinical sample. *Cochlear Implants Int.* 11(Suppl. 1), 355–359. doi: 10.1179/146701010X12671177989877

Power, R. J. D., and Dal Martello, F. D. (1997). From 834 to eighty thirty four: The reading of arabic numerals by seven-year-old children. *Math. Cogn.* 3, 63–85. doi: 10.1080/135467997387489

Semushina, N., and Mayberry, R. I. (2019). Numeral incorporation in Russian Sign Language. *Sign Language Stud.* 20, 83–131. doi: 10.1353/sls.2019.0015

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., et al. (2013). A brain area for visual numerals. *J. Neurosci.* 33, 6709–6715. doi: 10.1523/JNEUROSCI.4558-12.2013

Starrfelt, R., and Behrmann, M. (2011). Number reading in pure alexia: a review. *Neuropsychologia* 49, 2283–2298. doi: 10.1016/j.neuropsychologia.2011.04.028

Starrfelt, R., Habekost, T., and Gerlach, C. (2010). Visual processing in pure alexia: a case study. *Cortex* 46, 242–255. doi: 10.1016/j.cortex.2009.03.013

Taub, S. F. (2001). *Language From the Body: Iconicity and Metaphor in American Sign Language*. Washington DC: Cambridge University Press. doi: 10.1017/CBO9780511509629

Temple, P. A. (1989). Digit dyslexia: a category-specific disorder in development dyscalculia. *Cogn. Neuropsychol.* 6, 93–116. doi: 10.1080/02643298908253287

Titus, J. C. (1995). The concept of fractional number among deaf and hard of hearing students. *Am. Ann. Deaf*. 140, 255–262. doi: 10.1353/aad.2012.0582

Traxler, C. B. (2000). The Stanford Achievement Test, 9th Edition: national norming and performance standards for deaf and hard-of-hearing students. *J. Deaf Stud. Deaf Educ.* 5, 337–348. doi: 10.1093/deafed/5.4.337

Wollman, D. C. (1964). The attainments in english and arithmetic of secondary school pupils with impaired hearing. *Br. J. Educ. Psychol.* 34, 268–274. doi: 10.1111/j.2044-8279.1964.tb00636.x

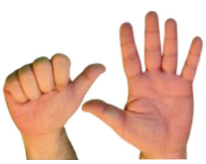
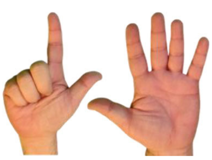
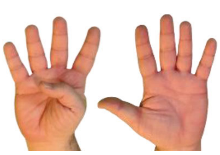
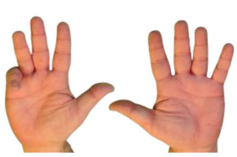
Wood, H. A., Wood, D. J., Kingsmill, M. C., French, J. R., and Howarth, S. P. (1984). The mathematical achievements of deaf children from different educational environments. *Br. J. Educ. Psychol.* 54, 254–264. doi: 10.1111/j.2044-8279.1984.tb02589.x

Zeshan, U., Ernesto, C., Delgado, E., Dikyuva, H., Panda, S., and De Vos, C. (2013). Cardinal numerals in rural sign languages: Approaching cross-modal typology. *Linguistic Typol.* 17, 357–396. doi: 10.1515/lity-2013-0019

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.

A. APPENDIX

TABLE A1 | Variants of digit-signs in ISL.

	Variant 1	Variant 2	Variant 3
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			

frontiers in Education

ORIGINAL RESEARCH
published: 30 June 2021
doi: 10.3389/feduc.2021.629053

Non-Symbolic Numerosity and Symbolic Numbers are not Processed Intuitively in Children: Evidence From an Event-Related Potential Study

Anne H. van Hoogmoed[1,2,3]*, Marije D. E. Huijsmans[3] and Evelyn H. Kroesbergen[1,3]

[1]Department of Educational and Learning Sciences, Utrecht University, Utrecht, Netherlands, [2]Nieuwenhuis Institute, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, Netherlands, [3]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

OPEN ACCESS

Edited by:
Korbinian Moeller,
Loughborough University,
United Kingdom

Reviewed by:
Mathieu Guillaume,
Stanford University, United States
Marie Coppola,
University of Connecticut,
United States

Reviewed by:
Madeline Quam,
University of Connecticut, in
collaboration with reviewer MC

***Correspondence:**
Anne H. van Hoogmoed
Anne.vanhoogmoed@ru.nl

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

Received: 13 November 2020
Accepted: 16 June 2021
Published: 30 June 2021

Citation:
van Hoogmoed AH, Huijsmans MDE and Kroesbergen EH (2021) Non-Symbolic Numerosity and Symbolic Numbers are not Processed Intuitively in Children: Evidence From an Event-Related Potential Study. Front. Educ. 6:629053. doi: 10.3389/feduc.2021.629053

The approximate number system (ANS) theory and the ANS mapping account have been the most prominent theories on non-symbolic numerosity processing and symbolic number processing respectively, over the last 20 years. Recently, there is a growing debate about these theories, mainly based on research in adults. However, whether the ANS theory and ANS mapping account explain the processing of non-symbolic numerosity and symbolic number in childhood has received little attention. In the current ERP study, we first examined whether non-symbolic numerosity processing in 9-to-12-year-old children ($N = 34$) is intuitive, as proposed by the ANS theory. Second, we examined whether symbolic number processing is rooted in non-symbolic numerosity processing, as proposed the ANS mapping account. ERPs were measured during four same-different match-to-sample tasks with non-symbolic numerosities, symbolic numbers, and combinations of both. We found no evidence for intuitive processing of non-symbolic numerosity. Instead, children processed the visual features of non-symbolic stimuli more automatically than the numerosity itself. Moreover, children do not seem to automatically activate non-symbolic numerosity when processing symbolic numbers. These results challenge the ANS theory and ANS mapping account in 9-to-12-year-old children.

Keywords: ANS mapping account, ANS theory, children, non-symbolic numerosity processing, symbolic number processing, ERP

HIGHLIGHTS

- Children's non-symbolic (NS) numerosity and symbolic number processing was assessed
- ERPs show that NS numerosity and symbolic number processing is not intuitive
- Instead, children process visual features of NS stimuli automatically
- The data do not support automatic activation of numerosity during number processing
- Thus, the results challenge the ANS theory and ANS mapping account in children

INTRODUCTION

Numerical processing is an important early marker of mathematical performance (e.g., Schneider et al., 2017). Numerical processing can be subdivided into non-symbolic *numerosity* processing (e.g., comparison between two sets of dots) and symbolic *number* processing (e.g., comparison between two Arabic numerals or number words). A prominent theory on non-symbolic numerosity processing is the ANS (approximate number system) theory. This theory states that approximate numerosity, i.e., the number of objects in a set, is intuitively extracted when one is confronted with a set of objects, such as a dot pattern (Dehaene, 1997). This means that the visual properties of a set of objects are removed or normalized, such that the numerosity of the set can easily be established, and that this process goes without much effort. The ANS mapping account concerns symbolic number processing, and theorizes that symbolic number processing in adults is rooted in non-symbolic numerosity processing. Approximate non-symbolic numerosity is thought to be activated automatically when processing symbolic numbers (Dehaene, 1997). There is currently a hot debate about whether the ANS theory and ANS mapping account hold or whether alternative theories are more likely to explain non-symbolic numerosity processing and symbolic number processing (see for example Leibovich et al., 2017, including commentaries on this paper). The aim of the present study was to examine whether the ANS theory and ANS mapping account do underlie non-symbolic numerosity processing and symbolic number processing in children. An event-related potential (ERP)-paradigm was employed to gain insight into the processing of non-symbolic numerosity and symbolic number.

ERP-research on the validity of the ANS theory and ANS mapping account in children is limited. ERP-research in adults shows both evidence confirming the ANS theory (Temple and Posner, 1998; Paulsen and Neville, 2008; Hyde and Spelke, 2009; Hyde and Spelke, 2012; Park et al., 2017; Van Rinsveld et al., 2020) and ANS mapping account (Dehaene, 1996; Temple and Posner, 1998; Pinel et al., 2001; Libertus et al., 2007), as well as more recent evidence against the ANS theory (Gebuis and Reynvoet, 2012; Soltész and Szűcs, 2014; Van Hoogmoed and Kroesbergen, 2018) and ANS mapping account (Van Hoogmoed and Kroesbergen, 2018). Children's numerical processing mechanisms may either be the same or different from those in adults. Research has shown that even infants seem to have a rudimentary understanding of non-symbolic numerosity (e.g., Xu and Spelke, 2000; Xu et al., 2005; Xu and Arriaga, 2007). However, it is not yet completely clear whether this is purely based on numerosity, or whether it is based on the visual features of a set of objects (Gebuis et al., 2016). Moreover, the development of symbolic number processing only starts at a later age.

From a developmental perspective, the early acquisition of symbolic number in young children may be intertwined with non-symbolic numerosity processing, as children usually start grasping numerical information by counting small amounts of non-symbolic objects (e.g., toys, pieces of fruit, or various body parts; Gelman, and Gallistel, 1978). Based on the ANS mapping account, this symbolic information will remain to activate their non-symbolic counterparts, even into adulthood. However, Carey (2004), Carey (2009), Carey (2011) claims that only small numbers are acquired based on non-symbolic numerosities, not based on the ANS, but on parallel individuation. The acquisition of numbers larger than four would be dependent on verbal counting routines and the notion that the next number in the routine represents N + 1 instead of direct mapping onto non-symbolic numerosities. However, other possible mechanisms that help children acquire number symbols have been proposed in several commentaries on Carey's paper (2011). For example, children's understanding of number is argued to occur prior to learning verbal principles such as counting, and this knowledge of number might foster development of numerical representations (Gelman, 2011; Gentner and Simms, 2011; Landy et al., 2011; Spelke, 2011).

In older children in kindergarten, symbolic number processing has been shown to be related to children's mapping skills (i.e., linking symbolic numbers and non-symbolic numerosities; Kolkman et al., 2013). It might thus be the case that symbolic number and non-symbolic numerosity processing in (young) children—in contrast to adults—(partly) rely on a common mechanism. However, this does not necessarily mean that children automatically activate numerosity when confronted with numbers as proposed by the ANS mapping account, especially not when they are older and more proficient in dealing with symbolic numbers. There is indeed evidence that processing of symbolic (large) numbers predicts processing of non-symbolic numerosity in kindergartners instead of vice versa, which suggests that symbolic processing does not necessarily build on the ANS (Lyons et al., 2018). Instead, there may be a bidirectional relationship between the development of symbolic and non-symbolic processing (Goffin and Ansari, 2019). Together, this implies that the processing of non-symbolic numerosity might not be as intuitive in children as assumed by the ANS theory and that the ANS may not be automatically activated when processing symbolic number, as proposed by the ANS mapping account.

The present study had two aims. The first aim was to investigate whether non-symbolic numerosity processing in children between 9 and 12 years of age is intuitive, in line with the ANS theory, or whether numerosities are processed based on the processing of visual features instead, as is proposed by alternative theories such as the sensory-integration theory (Gebuis et al., 2016) and sense of magnitude theory (Leibovich et al., 2017). Second, we examined whether children's processing of symbolic number can be explained by the ANS mapping account, or whether this processing is independent of numerosity, based on symbol-symbol associations (e.g., Reynvoet and Sasanguie, 2016).

Non-Symbolic Numerosity Processing

The ANS theory states that non-symbolic numerosity processing relies on an innate approximate number system (Dehaene, 1997). Non-symbolic stimuli are thought to be processed by an intuitive estimation of numerosity (i.e., the number of objects in a set), independently of physical features of the stimuli, such as the size of the objects. Proof of concept for this theory is mainly based on behavioral ratio effects within comparison tasks: Comparing two non-symbolic numerosities is more difficult (i.e., lower accuracy and slower reaction times) when these numerosities are closer in magnitude, and thus have a ratio closer to 1 (see Guillaume and Van Rinsveld, 2018 for a meta-analysis). This ratio effect is assumed to result from a mental number line wherein numerosities that are spatially located together are automatically co-activated, suggesting that non-symbolic numerosities are processed intuitively. Results from ERP research mirror the behavioral results by showing early ratio-dependent ERP amplitudes around 200 ms after stimulus presentation, suggesting that numerosity processing is fast (Temple and Posner, 1998; Libertus et al., 2007; Paulsen and Neville, 2008; Hyde and Spelke, 2009; Hyde and Spelke, 2011; Hyde and Spelke, 2012).

Although the ANS theory suggests that stimuli are processed independent of physical properties, physical features are inherently related to numerosity in real life. For instance, if one child has two pieces of candy and another child has four pieces of candy, then the second child's candy will occupy more of the visual space (i.e., total area and surface of the candy). According to the ANS theory, these visual features would be removed in a very early stage of numerical processing (e.g., Dehaene, 1997), after which numerosities are estimated or compared. However, instead of estimating numerosity after removal of visual features, one might better use visual properties of the objects to determine which child has the most candy.

To prevent the use of visual properties to estimate or compare numerosities, most research on non-symbolic numerosity processing therefore aims to control for visual properties of the stimuli. Even when using this kind of control, some studies still find early effects of numerosity, independent of visual features (Park et al., 2016; Fornaciai et al., 2017), which could be interpreted as evidence for the ANS theory. However, there are also studies that show that with proper control over visual features, effects of numerosity are absent or only starting around 650 ms (Gebuis and Reynvoet, 2012; Soltész and Szűcs, 2014; Van Hoogmoed and Kroesbergen, 2018). These results do not align with the ANS theory, since intuitive processing is unlikely to take such a long time, because it should take little effort. Hence, these results are better explained by alternative theories, such as the sensory-integration theory which posits that the integration of visual features is at the basis of an approximation of numerosity (Gebuis et al., 2016).

In children, it has become evident that the processing of non-symbolic stimuli relies more and more on actual non-symbolic numerosity with age and education, whereas physical properties of the stimuli become less relevant (Park, 2018; Piazza et al., 2018). This may reflect the increasing precision of the ANS (Halberda and Feigenson, 2008). Alternatively, it may reflect a growth in inhibition, withdrawing the child from intuitively responding to visual features and basing their decisions on the number of elements instead (Fuhs and McNeil, 2013; Gilmore et al., 2013). The argument of increasing precision of the ANS would result in early effects of numerosity (more specifically ratio) in the ERP with smaller and relatively short-lasting effects for visual properties. However, growth in inhibition would result in late effects of numerosity in the ERP in combination with early and possibly longer-lasting effects of visual properties. Our first aim was thus to examine whether non-symbolic numerosity processing is indeed intuitive, as proposed by the ANS theory (Dehaene, 1997), resulting in early effects of numerosity. Alternatively, children could process visual features more automatically than numerosity, which would be more in line with the sensory-integration theory (Gebuis et al., 2016), resulting in early effects of visual features in combination with later or no effects of numerosity.

Event-Related Potential Correlates of Non-symbolic Numerosity Processing in Children

Previous ERP research shows similarities in ERPs of non-symbolic numerosity processing between 5- and 8-year-old children and young adults (Temple and Posner, 1998; Heine et al., 2013; Soltész and Szűcs, 2014). Children and adults show similar neural activation over the parietal cortices when processing non-symbolic numerosity. For instance, similar ratio effects were displayed in the early ERP components N1 and P2p for children and adults (Temple and Posner, 1998). However, visual properties of the stimuli were not controlled in this study and only small numerosities were included (1–9). Other research controlling visual properties showed systematic numerosity distance effects in typically developing children in second and third grade in the parietal regions between 280 and 600 ms (Heine et al., 2013). Effects were found for subitizing, counting and estimation. The fact that effects for non-symbolic numerosity processing in children are more compelling for later ERPs (when controlling for visual properties of the stimuli), seems to indicate that this is not an automatic, but a more conscious process. In the current study, early effects of ratio would support the ANS theory, whereas either late ERP components related to numerosity, or no components related to numerosity at all, in combination with early and longer-lasting effects of visual features, may align better with the sensory-integration theory (also depending on the processing of the visual features of the stimuli).

Symbolic Number Processing

The ANS mapping account theorizes that symbolic number processing is rooted in the processing of the corresponding non-symbolic numerosity (Dehaene, 1997). As such, when encountering a number, the corresponding numerosity is assumed to be automatically activated in adults. Evidence for this account is mainly based on similar ratio effects for symbolic numbers and non-symbolic numerosities, which was assumed to be due to similar overlapping representations of numerosities and numbers (Dehaene et al., 1990; Verguts and Van Opstal, 2005; Holloway and Ansari, 2008; Sasanguie et al., 2012; Sasanguie et al., 2013). The timing of these non-symbolic ratio effects and symbolic distance effects is also similar, as has been shown by ERP research (Dehaene, 1996; Temple and Posner, 1998; Libertus et al., 2007). Arguments for the ANS mapping account thus seem convincing.

However, recent research has challenged the ANS mapping account, by raising several theoretical concerns about important assumptions (e.g., is it an evolutionary system; Reynvoet and Sasanguie, 2016) and caveats (e.g., inconsistent findings; Gevers et al., 2016) in those theories. For example, ratio and distance effects have also been found in non-numerical comparison tasks such as ordering letters of the alphabet, which do not have overlapping representations (Van Opstal et al., 2008). This implies that the effects are likely task-related instead of numerosity-related. Based on these results one cannot conclude that numerosity and number share the same numerical representation. Recently, symbolic numbers have been suggested to be processed independently of numerosity (Lyons et al., 2012; Sasanguie et al., 2017). Moreover, we showed that adults do not automatically activate numerosities when processing symbolic numbers (Van Hoogmoed and Kroesbergen, 2018). Measuring EEG (ERPs) during four different match-to-sample tasks (i.e., including non-symbolic numerosities, symbolic numbers, and combinations of both), we demonstrated that processing a non-symbolic target is different when the target is preceded by a non-symbolic prime compared to being preceded by a symbolic prime. If one would assume that a symbolic prime automatically activates the corresponding non-symbolic numerosity, one would expect that the processing of the non-symbolic target would not differ based on whether it is preceded by a symbolic or non-symbolic prime. As such, these results suggest that even when a task requires mapping (e.g., comparison between a symbolic number and a non-symbolic numerosity), symbolic stimuli are not automatically mapped onto their corresponding non-symbolic numerosities in adults.

From a developmental perspective, it seems that symbolic number processing is intertwined with non-symbolic numerosity processing in (young) children. When children learn numbers, they learn them by mapping these onto numerosities. For example, many children start learning numbers by counting their (and others) body parts (e.g., how many eyes, how many arms, how many fingers do you have?). However, symbolic skills appear to take a more prominent place than non-symbolic skills in the development of mapping skills in four-to six-year-old children (Kolkman et al., 2013). Whereas non-symbolic skills are related to symbolic skills and mapping skills in the first year of kindergarten, the relation between non-symbolic and symbolic skills becomes insignificant in the second year. Moreover, research shows that symbolic processing predicts non-symbolic skills as soon as children have initial number understanding, instead of the other way around (Lyons et al., 2018). This suggests that if these skills are still related in older children, non-symbolic (i.e., numerosity) processing may not be the primary format as proposed by the ANS mapping account. Instead, (larger) symbolic numbers may be acquired by the successor function (i.e., the next number in the counting row is exactly one more than the previous number), and may be embedded in a semantic network of numbers instead of grounded in the ANS (Krajcsi et al., 2016; Krajcsi et al., 2018; Reynvoet and Sasanguie, 2016). This may explain why the relation between non-symbolic and symbolic number weakens with age (e.g., Kolkman et al., 2013).

In purely symbolic tasks, children from kindergarten to third grade, as well as children in sixth grade have been found to use digits' physical properties to determine their magnitude, rather than their numerical value in a same-different task. No distance effect was found for numerical value (Defever et al., 2012). In a mixed notation task–in which digits needed to be compared to non-symbolic numerosities–a distance effect was found, showing no development with age until the end of primary school (Defever et al., 2012). In contrast, other research on symbolic digit comparison and comparison of non-symbolic numerosities (controlled for physical properties) found that the sizes of the symbolic and non-symbolic distance effects both decreased between six and eight years of age. The researchers concluded that children's magnitude representations become more precise as they grow older (Holloway and Ansari, 2009). Whether the distance effect becomes more fine-tuned with age or not, it seems evident that an effect is present in children, even when controlling for visual properties of the non-symbolic stimuli. Therefore, it could be that in children symbolic number processing is rooted in non-symbolic numerosity processing, in line with the ANS mapping account and findings in younger children. This may especially be the case when numbers need to be related to numerosities, which may involve either activation of the non-symbolic numerosity based on the processing of the symbolic number, or the activation of notation-independent code that is also activated by non-symbolic numerosities (Piazza et al., 2007). However, based on adult literature, it may also be that older children do not activate the corresponding numerosity in a purely symbolic task (e.g., Marinova et al., 2018; Marinova et al., 2021). In mapping tasks, they may map numerosities onto numbers, thus in the opposite direction as predicted based on the ANS mapping account (e.g., Van Hoogmoed and Kroesbergen, 2018).

Event-Related Potential Correlates of Symbolic Number Processing

Previous research shows several differences in ERP correlates of symbolic number processing between adults and children or adolescents (Temple and Posner, 1998; Soltész, Szűcs et al., 2007). While amplitude and direction of the P2p effect for ratio in a number comparison task in five-year-old children was similar to the effect of adults, the effect was delayed in children (Temple and Posner, 1998). Regarding the ratio effect, differences between adolescents (with math problems, and matched controls) and adults (without math problems) were found as well (Soltész et al., 2007). Slope and topography of the (late) ratio effects were different, being more mature in adults. These findings seem to indicate that symbolic processing changes over development, as differences in ERP components or scalp locations were found respectively. Soltész et al. (2007) proposed that differences in the late ERP components reflect differences in complex symbolic number processing between adolescents and adults. Note that similar differences in symbolic number processing between children and adults have been found in fMRI research (Ansari et al., 2005). These studies seem to suggest that numerical processing mechanisms in children and adolescents differ from adult mechanisms. This may indicate that symbolic number processing relates to non-symbolic numerosity processing in a different way in adults and children.

The Current Study

In the current study, we examined the electrophysiological correlates of the processing of non-symbolic numerosity and symbolic number in children between 9 and 12 years of age. Electro-encephalograms (EEGs) were administered during four match-to-sample tasks with two ratios, measuring the ratio effect in processing of non-symbolic stimuli, symbolic stimuli, or a combination between non-symbolic primes and symbolic targets and vice versa. The non-symbolic stimuli were controlled for visual features using the script of Gebuis and Reynvoet (2011). In this setup, ratio effects for non-symbolic processing, symbolic processing, and mapping of non-symbolic numerosities and symbolic numbers could be measured. Moreover, effects of task format and of visual properties (i.e., surface, area, and diameter) of the stimuli could be examined.

Based on the intuitive processing of non-symbolic numerosity as proposed by the ANS theory, one would expect early effects of ratio in the non-symbolic task. Moreover, one would expect no later or long-lasting effects of visual properties, since these would be removed/normalized in an initial step based on the ANS theory. However, based on the alternative sensory-integration theory, one would expect more long-lasting effects of visual properties, and only later or no ratio effects for numerosity. With regards to symbolic processing, based on the ANS mapping account, one would expect to find similar (possibly slightly delayed) ratio effects as compared to the non-symbolic condition, since numerosity would be automatically extracted from the symbolic stimuli, and then processed similarly to the non-symbolic stimuli (i.e., either in a non-symbolic or notation-independent format). An additional way to examine whether non-symbolic numerosity and symbolic number are mapped onto each other (based on the ANS theory), is a comparison between tasks with the same primes or targets. If symbolic numbers and non-symbolic numerosities are mapped onto each other, one would expect that the processing of non-symbolic primes is the same, regardless of whether these are followed by non-symbolic or symbolic targets. Similarly, the processing of non-symbolic targets would be the same regardless of whether these are proceeded by non-symbolic or symbolic primes. Thus, differences in processing of non-symbolic stimuli depending on task, would provide evidence against the ANS mapping account.

METHOD

Participants

Participants were 50 children from grade 3 to 6 in primary school. Seven children were excluded due to recording problems (partly due to a broken Ground-electrode), and nine children were excluded due to noisy data (see below). The final sample consisted of 34 children (16 boys and 18 girls) in grade 3 to 6, with a mean age of 10.71 years (SD = 0.78). All participants had normal or corrected-to-normal vision.

Procedure

Participants were recruited via a letter they received from their schools. The major part of the participants (24 participants) were tested individually in a separate room within their schools. The procedure was explained to them verbally with supporting pictures. The other part of the participants (10 participants) was tested in the lab. The children were informed that they could choose to stop participating at every moment. Informed consent was signed by their parents.

After applying the EEG, the participants were seated behind the computer. The task instruction was read from the screen together with the child. Participants were told that there would be a break after each task. The administrator of the task stayed with the child during the experiment to answer any questions, and if necessary, to encourage the child to pursue. The ERP tasks were presented in a random order. After the four tasks, the EEG cap was removed from the participant. All tasks including application and removal of the EEG-cap lasted about 45–60 min. The parents of all the participants gave written informed consent in accordance with the Declaration of Helsinki. The research was approved by the ethics review board of the faculty of Social and Behavioral Sciences of Utrecht University.

Tasks

Non-Symbolic (Ns-Ns)

In the non-symbolic task, trials consisted of a prime picture with dot patterns and a target picture with dot patterns, see **Figure 1**. The dot patterns were generated in Matlab with the script described in Gebuis and Reynvoet, 2011. This script allows for controlling the relation between the number distance and visual properties, as well as the congruency in area subtended, density, total surface of the dots, average diameter, and total circumference. The visual properties of the stimuli were documented, enabling division of the data based on visual properties as well (Gebuis and Reynvoet, 2011). The number of dots for the primes ranged between 20 and 40, with both smaller and larger targets at ratio 0.5, and 0.7. As such, all numbers ranged between 10 and 80. A trials started with the presentation of a prime for 750 ms, then a blank screen jittered between 400 and 600 ms, and a target presented for 750 ms. The inter trial interval was jittered between 1,000 and 1,500 ms. In total, 88 trials were presented to the children. Twenty trials were presented for each distance x size (target larger vs. target smaller than prime). In addition to that, we included eight trials in which the numerosity of the prime and the target were the same. Participants were instructed to passively watch the stimuli and only respond by pressing the space bar if they thought the prime and target stimuli displayed the same quantity as soon as possible during stimulus presentation or during the following blank screen (ITI).

Non-Symbolic–Symbolic (Ns-S)

The Ns-S task was identical to the Ns-Ns task with the exception that the targets were presented as digits instead of dot patterns. Moreover, both the prime and target were presented for 1,000 ms.

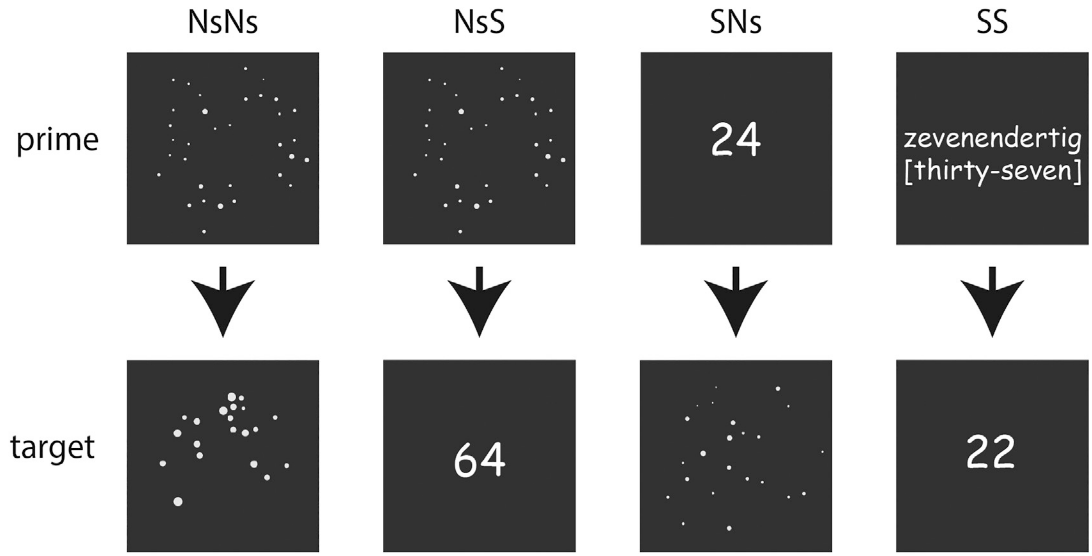


FIGURE 1 | Overview of the stimuli-formats in the different tasks with the upper line presenting the primes for each task, and the lower line presenting the targets.

Symbolic–Non-symbolic (S-Ns)

The S-Ns task was identical to the Ns-S task with the exception that the primes were presented as digits instead of dot patterns and targets were presented as dot patterns instead of digits.

Symbolic (S-S)

The S-S task was identical to the Ns-S task with the exception that the primes were presented as number words, instead of dot patterns.

Analyses

Behavioral

Participants were instructed to only respond to trials in which the prime and target depicted the same numerosity. As such, a non-response to the trials in which the prime and target did not match was taken as correct response. Mean accuracy per ratio, per task was calculated in SPSS, version 23. To examine whether performance was above chance in each task, one-sample t-tests were carried out against a test-value of 0.5. Bivariate correlations between performance and age were carried out separately for each task.

Event-Related Potential

Recording and Preprocessing

For the 24 participants tested at schools, data were recorded with a 32 electrode active cap (Biosemi, Amsterdam, Netherlands) with a sampling rate of 2048 Hz. Additional electrodes were placed on both mastoids, and below and next to the eyes. The system records data without reference. The electrode offset was kept below 50 μV. For the ten participants that were tested in the lab, data were recorded with a 32-electrode ActiCAP (Brain Products GmbH) and were recorded online with a sampling rate of 500 Hz. Measured activity was filtered online using a 125 Hz lowpass filter, and a time constant of 10 s. Impedance was kept below 50 μV.

After recording, all data were imported into Matlab 2017a (The MathWorks Inc., 2017) and analyzed using the Fieldtrip toolbox (Oostenveld et al., 2011). Data of all participants were downsampled to 500 Hz, rereferenced to the linked mastoids, and low-pass filtered at 40 Hz. ICA was used to identify and delete eye blinks and horizontal eye movements. After that, data were manually inspected for bad channels. Bad channels were removed and replaced with a weighted sum of the surrounding channels. Deleted channels were never adjacent to each other. Data (primes and targets) were segmented from 200 ms before to 1,000 ms after stimulus onset and baseline corrected. After artifact rejection, the data were averaged per ratio per task for the targets (M_{Ntrials} = 34.2, range 24–40). Data from target larger than prime and target smaller than prime were collapsed because of the limited number of trials included (Van Hoogmoed and Kroesbergen, 2018). To examine the effects of visual properties, averages were generated for small and large diameter, small and large area, and small and large surface based on all non-symbolic primes and targets (mean Ntrials = 31.7, range N_{trials} 19–40). The 40 trials with the largest and 40 trials with the smallest surface/area/diameter were selected to generate averages with similar amounts of trials as compared to the number of trials per ratio. Averages for primes and targets within each task were also computed (M_{Ntrials} = 71.9, range 53–88).

Analyses

Grand averages were computed over the 28 common electrodes in both recording systems. Since the time course of the differences between conditions was unknown because of the mixed findings in previous research, cluster based permutation tests were carried out (Oostenveld et al., 2011). For the Ratio effects in the tasks,

TABLE 1 | Mean proportion correct and standard deviations for matching and non-matching trials per task.

	Matching M (SD)	Matching RT (SD)	Non-matching ratio 0.5 M (SD)	Non-matching ratio 0.7 M (SD)
NsNs	0.36 (0.22)	598.7 (86.9)	0.76 (0.15)	0.69 (015)
NsS	0.51 (0.19)	700 (99.7)	0.76 (0.15)	0.62 (0.14)
SNs	0.34 (0.21)	834.3 (247.3)	0.68 (0.14)	0.64 (0.14)
SS	0.80 (0.21)	599.0 (85.5)	0.99 (0.05)	0.99 (0.02)

NsNs = task with non-symbolic primes and non-symbolic targets, NsS = task with non-symbolic primes and symbolic targets, SNs = task with symbolic primes and non-symbolic targets, and SS = task with symbolic primes and symbolic targets.

four separate permutation tests were carried out, one for each task. Similar permutation tests were performed for the physical parameters on small vs. large (mean) diameter, area (within the convex hull), and surface (of the dots). To test for differences in the processing of non-symbolic and symbolic stimuli depending on task, three cluster based permutation tests were carried out: one to compare the processing of non-symbolic primes in the NsNs-task vs. the NsS-task, one to compare the processing of the non-symbolic targets in the NsNs-task vs. the SNs-task, and one to compare the processing of the symbolic targets in the NsS-task and the SS-task.

A dependent-samples *t*-test on amplitude for each channel x sample between 0 and 1,000 ms served as input for the cluster based permutation test. Spatio-temporal clusters were defined based on these t-statistics ($\alpha = 0.05$). These clusters were entered into the cluster-based permutation test (500 permutations or 1,000 permutations if the obtained $|p\text{-}\alpha| < 0.002$). Since cluster-based statistics (clusterstats) are calculated for positive and negative slopes separately, the *p*-values were compared to $\alpha = 0.025$ (0.05/2) for all analyses (Fieldtrip, n. d.; for an example see Van Hoogmoed and Kroesbergen, 2018). Note that by using cluster-based permutation tests, the whole cluster is tested as one test-statistic. As such, the latency and exact location of a cluster are only descriptive (Sassenhagen and Draschkow, 2019).

RESULTS

Behavioral

Accuracy for each task is reported in **Table 1** for matching pairs and non-matching pairs separately. For matching pairs, a correct response was a button press, whereas for non-matching pairs, a correct response consisted of refraining from a button press. Performance was above chance on task level in all tasks; for the NsNs task ($M = 0.69$, $SD = 0.12$), $t\,(33) = 9.74$, $p < 0.001$, $d = 1.67$, for the NsS task ($M = 0.67$, $SD = 0.12$), $t\,(33) = 8.67$, $p < 0.001$, and $d = 1.49$, for the SNs task ($M = 0.63$, $SD = 0.11$), $t\,(33) = 6.74$, $p < 0.001$, $d = 1.16$, and for the SS task ($M = 0.97$, $SD = 0.03$), $t = 86.13$, $p < 0.001$, and $d = 14.77$. However, participants had difficulty identifying matching trials, with the proportion of correctly answered trials ranging from 0.34 to 0.80. Age of the participants did not significantly relate to performance in any of the tasks ($0.038 \le r \le 0.255$). Exploratory analyses revealed a small difference between the performance on the NsS-task and the SNs-task, $t\,(33) = 2{,}12$, $p = 0.041$, $d = 0.36$.

Event-Related Potential

Distance Effects for Ratio per Task

ERPs for the different Ratios are depicted in **Figure 2** for each task separately. A permutation test on the difference between ratio 0.5 and ratio 0.7 in the NsNs-task resulted in a significant negative cluster for Ratio, largest negative clusterstat = −1836.8, $p = 0.022$, reflecting a parietal distance effect between 650 and 950 ms (see **Figure 3**). No significant positive cluster was found, largest positive clusterstat = 148.3, $p = 0.329$. For the NsS-task, no positive clusters were found. Moreover, no significant negative cluster was found, largest negative clusterstat = −352.2, $p = 0.216$, which reflects no significant ratio effect for the NsS-task. For the SNs-task, no significant positive or negative clusters were found, largest positive clusterstat = 161.1, $p = 0.371$, largest negative clusterstat = −100.8, $p = 0.485$, which reflects no significant ratio effect for the SNs-task. For the SS-task, two significant effects of distance were found. The first cluster, positive clusterstat = 2121.9, $p = 0.012$, reflects a very early effect, around 50–300 ms over the centro-parietal scalp regions. The second cluster, positive clusterstat = 2089.2, $p = 0.012$, reflects a broadly distributed effect between 700 and 800 ms (See **Figure 4**). No negative clusters were found.

Distance Effects for Visuals

ERPs for the different visual features are depicted in **Figure 5**. For the visual feature area, the permutation test (1,000 permutations) resulted in no significant positive or negative clusters, largest positive clusterstat = 223.7, $p = 0.241$, largest negative clusterstat = −1217.3, $p = 0.035$. For the visual feature surface (1,000 permutations), the results showed no significant positive cluster, largest clusterstat = 333.7, $p = 0.208$, but a significant negative cluster, clusterstat = -2499.0, $p = 0.013$. This cluster reflects a relatively early difference from around 200 ms to around 300 ms over the parieto-occipital scalp regions (see **Figure 6**). For the visual feature diameter, the results showed no significant positive cluster, largest clusterstat = 275.8, $p = 0.275$, but a significant negative cluster, clusterstat = −2005.3, $p = 0.009$, reflecting a occipital difference between small and large diameter from around 200–300 ms (see **Figure 7**).

Differences Between Tasks

To assess whether processing of stimuli is related to task format, we investigated differences in processing non-symbolic primes in the NsNs-task and NsS-task, non-symbolic targets in the NsNs-task and SNs-task, and symbolic targets in the NsS-task and SS-task (see **Figure 8**).

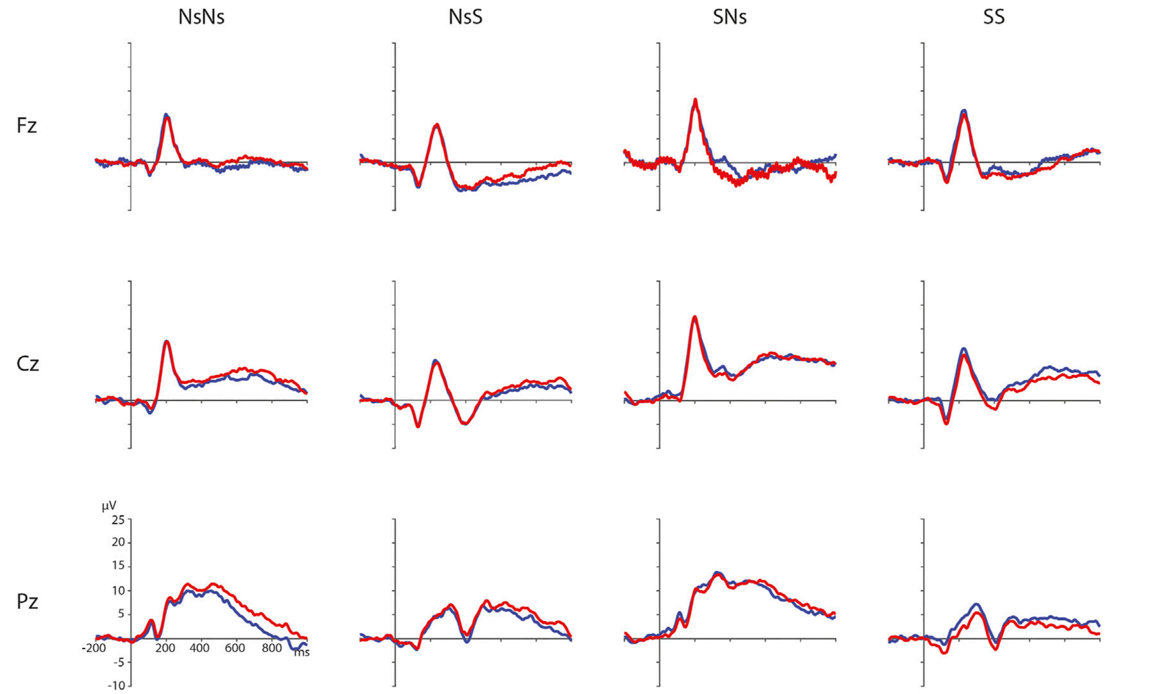


FIGURE 2 | Distance effects per task on electrodes Fz, Cz, and Pz (electrodes chosen for illustration purposes only) with ratio 0.5 in blue and ratio 0.7 in red.

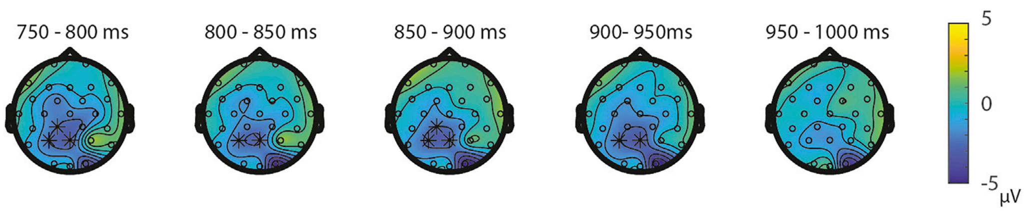


FIGURE 3 | Topoplots of the differences between the ratio's in the NsNs-task per time window with asterisks representing the significant differences between the ratios based on α = 0.05.

Symbolic primes were not compared between tasks, since the primes in the SNs-task consisted of digits, whereas the primes in the SS-task consisted of written number words. For the non-symbolic primes, no significant differences were found between the NsNs-task and the NsS-task, largest positive clusterstat = 218.1, $p = 0.307$, largest negative clusterstat = −1148.2, $p = 0.054$, showing no evidence for differences in the processing of non-symbolic primes based on whether it is followed by a symbolic target or by a non-symbolic target. For the non-symbolic targets, no significant positive clusters were found, largest positive clusterstat = 392.2, $p = 0.152$, but a significant negative cluster was found, clusterstat = −6507.1, $p = 0.012$. This cluster reflects a widespread (fronto)central difference moving toward the parietal scalp regions between 100 ms 350 ms (see **Figure 9**) reflecting a difference between the processing of non-symbolic targets in the NsNs-task and the NsS-task. For the symbolic targets, a significant positive cluster was found, clusterstat = 3724.4, $p = 0.012$, reflecting a long lasting occipital difference (200–700 ms) between the processing of symbolic targets in the NsS-task and the SS-task (see **Figure 10**). No significant negative clusters were found, largest clusterstat = −2141.2, $p = 0.040$.

DISCUSSION

The first aim of the present study was to examine whether non-symbolic numerosity processing in 9-to-12-year-old children is intuitive and thus relatively fast, in line with the ANS theory (Dehaene, 1997), or whether visual properties of stimuli play a role in processing the numerosity, in line with the sensory-integration theory (Gebuis et al., 2016; Gevers et al., 2016). The second aim was to examine whether children process symbolic numbers by automatically mapping them onto non-symbolic numerosities, in line with the ANS mapping account

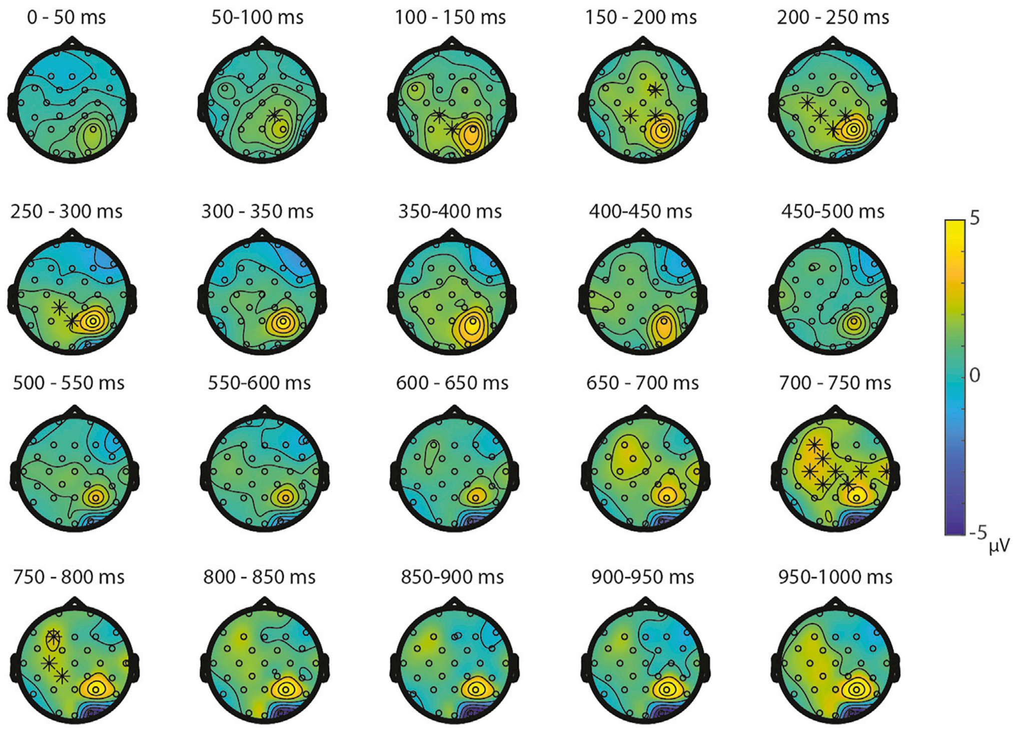


FIGURE 4 | Topoplots of the differences between the ratio's in the SS-task per time window with asterisks representing the significant differences between the ratios based on α = 0.05.

(Dehaene, 1997). Alternatively, children may process numbers without automatically activating the corresponding numerosity, for example based on symbol-symbol associations (Krajcsi et al., 2016; Reynvoet and Sasanguie, 2016).

Non-Symbolic Numerosity Processing

The behavioral data in the non-symbolic task show performance above chance level. However, accuracy on the matching trials was relatively low, which indicates that the children had difficulty determining whether two numerosities were the same. The accuracy on ratio 0.5 was higher than on ratio 0.7, which is in line with the expectations based on previous research (e.g., Guillaume and Van Rinsveld, 2018). The ERP data showed only a late parietal effect for ratio in the completely non-symbolic task, starting at 650 ms after the presentation of the target, in line with previous research (Soltész and Szűcs, 2014). However, earlier effects for the visual feature surface and diameter were found over the occipital scalp regions, starting around 200 ms, suggesting that these are processed automatically, even though the task focused on numerosity. Children thus seem to process visual features more automatically than numerosity of non-symbolic stimuli. Based on the ANS theory, one would have expected an earlier distance effect for ratio of numerosity. As such, our findings do not convincingly support the ANS theory (Dehaene, 1997).

In earlier research in adults, we argued that the presence of a long-lasting effect of visual properties starting early was not in line with the ANS theory (Van Hoogmoed and Kroesbergen, 2018). Instead, we argued that these findings were in line with two recent theories that have been proposed as alternatives for the ANS theory: The sensory-integration theory and sense of magnitude theory (Gebuis et al., 2016; Leibovich et al., 2017), both suggesting that visual features are not removed before processing numerosity, but are at the basis of this process. However, the results in adults differed from those in children. The duration of the effect of the visual features was shorter and a late ratio effect was present in children, but not in adults. As such, the results found in this study do not support the sensory-integration theory either. Another possible hypothesis may be that children do not use visual features as the basis for the processing of numerosity, but first process the visual features, and inhibit their response based on these visual features (Fuhs and McNeil, 2013; Gilmore et al., 2013). After that, they may still consciously process the numerosity, causing the late effects of ratio. However, this possibility requires additional research.

Another difference between the effects of visual properties in previous research with adults (Van Hoogmoed and Kroesbergen, 2018) and the current study in children was that the visual features showing a distance effect in the non-symbolic task in children were surface and diameter as opposed to area in adults. This difference, however, complements a developmental study on the effects of visual features in children and adults (Gilmore et al., 2016). That study revealed that adults indeed rely more on the convex hull (or area) of non-symbolic stimuli, whereas in primary school children the dot size (or surface) was most important. This

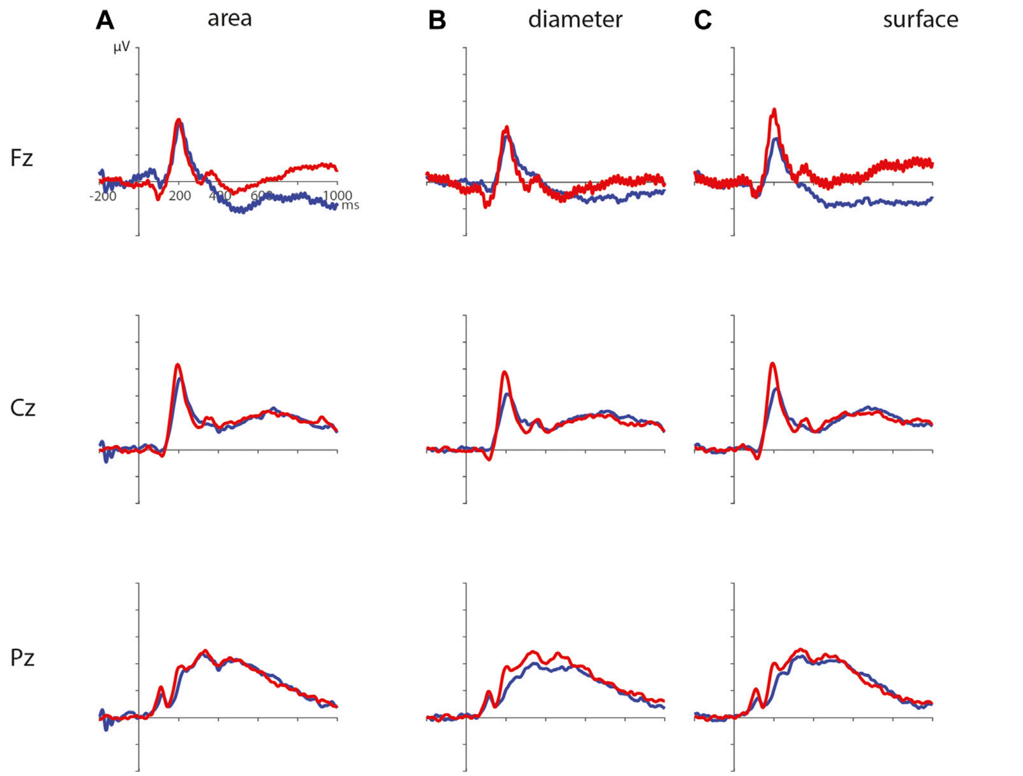


FIGURE 5 | ERPs for the different visual features with small area **(A)**, diameter **(B)**, and surface **(C)** in blue, and large area, diameter, and surface in red.

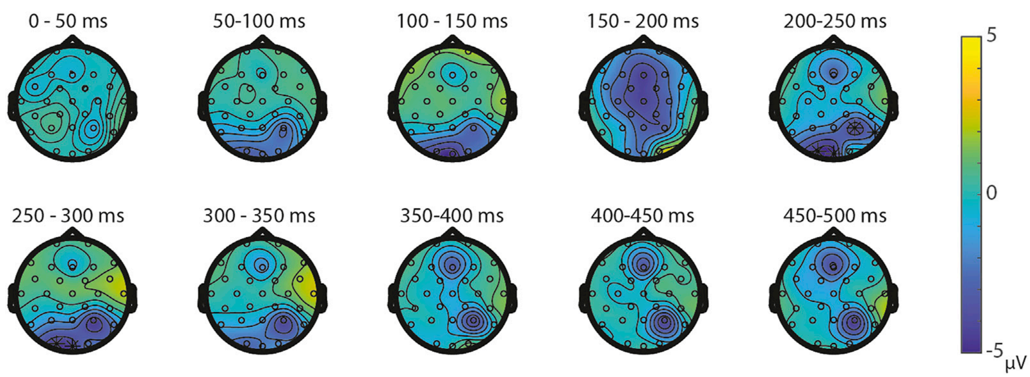


FIGURE 6 | Topoplots of the differences between small and large surface per time window with asterisks representing the significant differences between the small and large surface based on α = 0.05.

is in line with the current results. Additionally, it has been suggested that the contribution of any visual feature to non-symbolic numerosity processing is dependent on the type of stimuli, the setting, and the context of a task (Leibovich and Henik, 2014). As such, the specific visual feature that shows the result is deemed less important than the finding of effects for visual features itself.

Our results differ from the recent results on automatic processing of numerosity, even as early as the visual cortex (Park et al., 2016; Fornaciai et al., 2017; Starr et al., 2017; Park, 2018; DeWind et al., 2019). In these studies, a fixed number of five numerosities was used instead of the ratios of numerosities that were used in the present study. Moreover, control over visual properties was carried out in a different manner in which visual properties seem to correlate more with the presented numerosity compared to the current study. In such passive paradigms without focus on numerosity, it is likely that participants automatically process the most salient features of the stimuli (either numerosity or a visual property; see Leibovich et al., 2017). In our paradigm, the focus of the

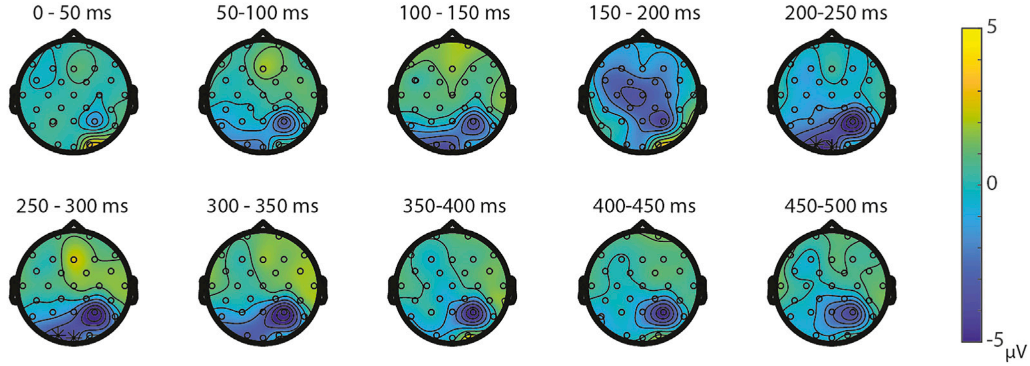


FIGURE 7 | Topoplots of the differences between small and large diameter per time window with asterisks representing the significant differences between the small and large diameter based on α = 0.05.

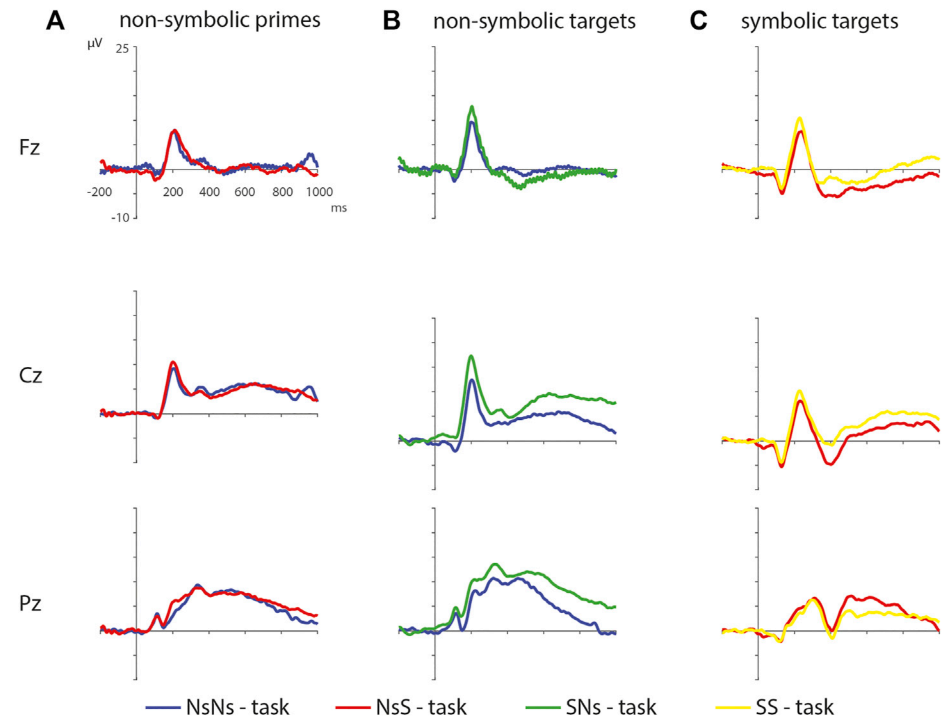


FIGURE 8 | ERPs of the non-symbolic primes in the NsNs-task and the NsS-task **(A)**, the non-symbolic targets in the NsNs-task and the SNs-task **(B)**, and the symbolic targets in the NsS-task and the SS-task **(C)**.

participants was specifically on numerosity, since we instructed them to press a button in case of matching numerosities. Even then, there was more evidence for early processing of visual features as compared to processing of the numerosity, which suggests that this was more intuitive.

To conclude, we argued that our results do not align with the ANS theory, but neither with the sensory integration theory. This argument was based mainly on the late effects of ratio, which are in contrast with earlier ERP-studies showing P2p-effects in comparison tasks (e.g., Temple and Posner; Heine et al., 2013). The current task differs from the comparison task in that it may first be necessary to estimate the numerosity (after removal/normalization of visual features), and then compare it to the numerosity that is kept in memory. The normalization of visual features may be reflected in the effects of visual features around 200–300 ms. The late parietal effect for ratio starting around 650 ms may reflect the next step, i.e., the comparison itself. As such, the process could be similar to what would be expected based on the ANS theory. However, the timing of the effects suggest that the

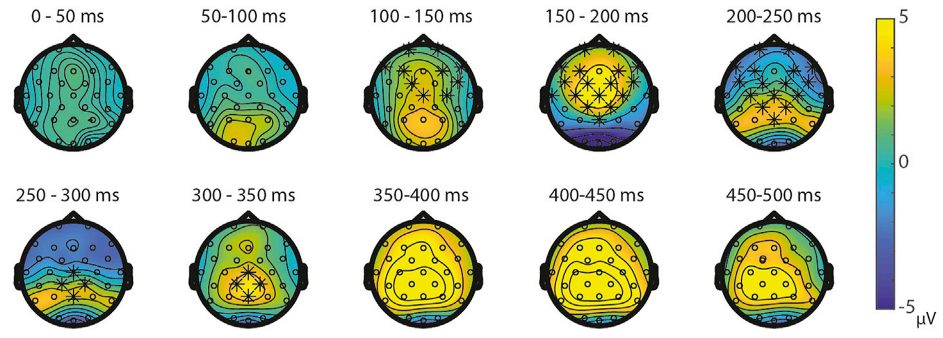


FIGURE 9 | Topoplots of the differences between the non-symbolic targets in the NsNs-task and the SNs-task with asterisks representing significant differences based on α = 0.05.

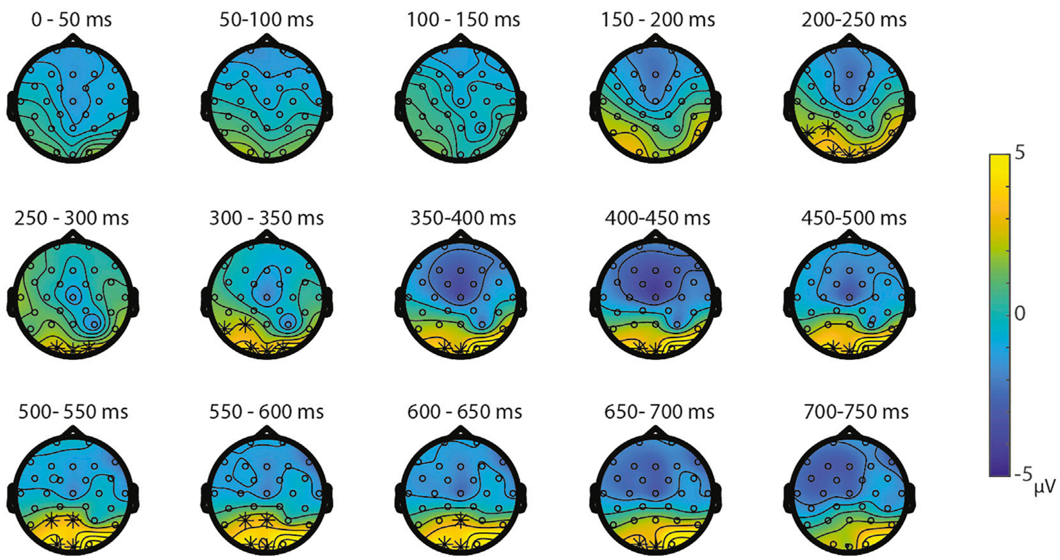


FIGURE 10 | Topoplots of the differences between the symbolic targets in the NsS-task and the SS-task with asterisks representing significant differences based on α = 0.05.

processing of the numerosity is not intuitive, but requires effort.

Symbolic Number Processing

For the completely symbolic task, we found a very early effect of ratio between 50 and 300 ms over the central scalp, and a late effect for ratio over the central scalp between around 700 and 800 ms. A ratio effect in a symbolic task may reflect overlapping representations of non-symbolic numerosity and symbolic number (Van Opstal and Verguts, 2011), which would support the ANS mapping account. However, this would mean that the symbolic numbers activate the non-symbolic representation and show a distance effect based on these non-symbolic representations. In that case, one would expect a similar ratio effect in the non-symbolic task as well, something that was not supported by our results. The early effect was not present in the non-symbolic task. Moreover, the late effect was spread over the (left)-fronto-central scalp in the symbolic task and more parietal in the non-symbolic task. The results thus suggest that non-symbolic numerosity processing and symbolic number processing rely on different mechanisms. This idea is supported by recent work showing only weak relations between the processing of non-symbolic numerosity and symbolic number in adults and children (see Leibovich and Ansari, 2016 and Reynvoet and Sasanguie, 2016 for reviews).

The ratio effect found in the symbolic task may be better explained by alternative accounts, such as the discrete semantic system (DSS; Krajcsi et al., 2016) or a symbol-symbol association account (Reynvoet and Sasanguie, 2016). Both accounts suggest that relations between symbolic numbers are at the basis of symbolic number processing instead of relations between a symbolic number and the corresponding numerosity. According the DSS theory, numbers are stored in nodes in a network, similar to the mental lexicon or other conceptual networks. The strength of the connections between the nodes is proportional to the strength of the semantic relations (Krajcsi et al., 2016). Related to this theory, Reynvoet and Sasanguie argue initially small numbers gain meaning through coupling to the OTS, but larger numbers gain meaning through the ordinal relation between numbers. The strength of connections between the nodes (stronger connection for numbers closer to each other) or the ordinal relation between numbers may explain the ratio effects found in symbolic tasks, including the symbolic task in the current study. With regards to the mapping tasks, based on the ANS mapping account, one would expect that humans automatically activate numerosity when they are processing symbolic numbers, especially when these need to be mapped onto non-symbolic numerosities. In this study, this would have resulted in similar ratio effects in all tasks, including the mapping tasks (although maybe different in exact timing). However, no significant effects of numerosity were found in the mapping tasks. Moreover, if numerosity would be activated automatically, this activation should be independent of task format. Thus, based on the ANS mapping account, non-symbolic numerosity processing should not differ between completely non-symbolic tasks and mapping tasks. The ERP data showed no evidence for differences in processing non-symbolic primes within different task formats, but do show significant differences in the processing of non-symbolic targets between task formats. This implies that the activation of non-symbolic numerosity is not automatic, and thus not in line with the ANS mapping account.

The differences in non-symbolic processing based on task format contradict fMRI studies showing that, under passive viewing conditions, non-symbolic and symbolic stimuli both activate the hIPS (Dehaene et al., 2003; Piazza et al., 2007). The authors interpreted these results as evidence for the integration between symbolic numbers and non-symbolic numerosities. Their findings thus seem to be in line with the ANS mapping account. However, the fact that the hIPS is involved in symbolic and non-symbolic processing does not necessarily mean that both formats are processed in the same manner. Moreover, more recent, fine-grained fMRI research shows different regions involved in symbolic vs. non-symbolic processing (e.g., Bulthé et al., 2014; Bulthé et al., 2015; Holloway, Price, and Ansari, 2010; see Sokolowski et al., 2017 for a meta-analysis), indicating that both formats may be processed in different ways. The effect in our study indeed reflected larger ERP amplitudes for non-symbolic targets following symbolic primes compared to non-symbolic primes. This may indicate that more resources are allocated to non-symbolic numerosity processing during mapping than during the comparison of similar formats (Kadosh, Lammertyn, and Izard, 2008; Landgraf et al., 2010). Thus, our results substantiate other research showing differences in non-symbolic processing based on task format.

These results raise the question based on which format mapping tasks are solved. In our previous study in adults (Van Hoogmoed and Kroesbergen, 2018), we argued that tasks with mixed stimuli might be solved by attaching a symbolic number to the numerosity of a dot pattern and comparing this symbolic number to the presented digit. As such, it was suggested that symbolic number processing was the primary format in mapping tasks. The exploratory analysis on the differences in behavior between the SNs-task and the NsS-task in the current study could be seen as support for this argument. These results show that it was easier for the children to compare a symbolic target to a non-symbolic prime than to compare a non-symbolic target to a symbolic prime. However, whereas these results are provide some evidence against the notation-independent code, they do not necessarily inform us about the primary format of processing. The results could imply that the non-symbolic prime already activated the symbolic number, after which the symbolic target was matched to this symbolic number, meaning that symbolic processing was the primary format. However, it could also mean that the format of processing is dependent on the format of the prime, and that activation of the non-symbolic format based on a symbolic target (in the NsS-task) was easier than activation of the symbolic format based on a non-symbolic target (in the SNs-task). Moreover, the ERP results showed differences in the processing of the symbolic targets based on task format as well. If number would be the primary format of processing, one would not expect a dependency on task. This may suggest that non-symbolic numerosity processing is not rooted in symbolic number processing either, but processing format may be based on specific task demands.

Additional insights would have been possible based on a comparison between symbolic primes in the different task formats. However, we could not compare the symbolic primes to subsequent non-symbolic targets and symbolic targets, because the symbolic primes we used had a different format in both tasks. The symbolic primes consisted of digits (e.g., "10") in the mapping task, whereas the symbolic primes consisted of number words (e.g., "ten") in the symbolic task. The reason for this is that we prioritized the analyses of the (symbolic) targets: In the completely symbolic task, the prime and target had to be physically different in order to prevent children to compare visual properties of the number instead of its magnitude. Still, the notion that children do not anticipate the format of the target seems to hold, as there was no difference between the non-symbolic primes. This appears to suggest that children remember the (non-symbolic) prime in the original format, and only start processing the prime when the target is presented. This may explain the difference between the ERPs to symbolic targets in the mapping task vs. the symbolic task, but does not inform us about the format of processing. The behavioral data suggest that the mapping task was more difficult for children, which probably resulted in the differences in ERPs. This delayed processing of the prime in children may well be different from the way adults approach the task, irrespective of differences in the processing of numerosities and numbers itself.

Limitations

Several limitations of the present study should be acknowledged. First, the absence of responses to the non-matching trials in the experiment might have confounded the results, because non-response trials (in this case 91% of the trials) might be due to irrelevant functions (e.g., distraction, boredom, uncertainty in decision-making) instead of correct task performance and may have made the task more difficult compared to an active comparison task. This difficulty can be inferred from the low accuracy on the matching trials. However, we explicitly aimed for a match-to-sample task with mainly non-response, because previous research on non-symbolic numerosity and symbolic number has suggested that active same-different tasks and active comparison tasks may tap into different cognitive processes (Van Opstal et al., 2008; Van Opstal and Verguts, 2011). Whereas the ratio effect in comparison tasks may be caused by a general decision process, in the same-different task, it is thought to be due to co-activation based on neural overlap between numbers. While acknowledging the limitations of the more passive task, the behavioral results show that there is a ratio effect in all tasks included non-symbolic stimuli. This suggests that children were actively engaged in the task, although their engagement may have been lower than in a traditional comparison task which requires a response to each trial.

A second limitation is the quite large age range of the participants. Previous research showed that susceptibility to perceptual cues is affected by the age of the participating children (Defever et al., 2013). However, the age range of participating children was much larger in that study. Moreover, our behavioral results did not show significant relations between performance and age. Therefore, we did not include age as a factor in our analyses. Future research with a larger sample and a larger age range could shed light on possible differences between different age groups.

A third limitation is that we did not directly examine the effect of visual properties on the processing of numerosity. Instead, we examined the effects of numerosity and visual features separately, and qualitatively compared them. Moreover, we chose to create stimuli such that the overlap between numerosity and visual features was as limited as possible (based on Gebuis and Reynvoet, 2011). This has led to a larger differences in visual features as compared to ratio differences. Previous research has shown that the variation in visual features impacts the judgment of numerosity (e.g., Nys and Content, 2012; Clayton et al., 2015; Gilmore et al., 2016). Future research should include a full comparison of the different methods of constructing visual stimuli (including more difficult ratios), and directly compare the impact of these different ways of constructing stimuli on numerosity processing. Such an endeavor would help to gain a comprehensive understanding of numerosity processing under different circumstances.

CONCLUSION

To conclude, our results show very late numerosity-related ratio effects, in combination with early effects related to the visual features of non-symbolic stimuli. As such, these results seem to contradict the ANS theory, and suggest that processing of non-symbolic numerosity is unlikely to be automatic. Moreover, we found that non-symbolic numerosity is not automatically activated when processing symbolic numbers, which contrasts the ANS mapping account (Dehaene, 1997). Although children can relate numbers and numerosities, given their behavioral ratio effect in the mapping tasks, this process does not seem to be automatic. In adults, it has been suggested that symbolic number could be the primary format of processing, and non-symbolic numerosity processing would possibly occur by estimating the number of dots, and then compare the numbers in a symbolic format (Van Hoogmoed and Kroesbergen, 2018). However, this hypothesis does not seem to hold either, since we found that the processing of symbolic targets is also dependent on task format in children. This may however be due to the fact that children, contrary to adults, do not anticipate on the upcoming target. The difference between the symbolic targets in the mapping task and the purely symbolic task might be explained by the notion that children still need to process the prime once the target it presented. Future research including both a blocked design and a mixed design (i.e., manipulating expectancy of a certain format) would be suitable to examine this idea. Moreover, future research including younger children could shed light on differences in the dependence or independence of symbolic number processing on non-symbolic numerosity processing over development. This may substantiate the current evidence against the ANS theory and ANS mapping account.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/a85rs/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Review Board of the Faculty of Social and Behavioral Sciences of Utrecht University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AvH and EK set up the experiment, AvH set up data collection and data analysis. AvH and MH analyzed the data. AvH, MH, and EK contributed to writing the manuscript.

FUNDING

This research was funded by the Dutch Scientific Organization (NWO), Aspasia grant number 015.008.028, awarded to EK.

ACKNOWLEDGMENTS

We would like to thank Dennis Hofman and Jos Jaspers for the technical support during data collection. We would like to thank Sanne van der Ven, Isabelle van Hasselt, Daniela van Wijngaarden, Anniek van Dijk, Jasmijn van Engeland, Hilde Vuijk-Biewenga, and Vivian Campschoer for their help in collecting the data.

REFERENCES

Ansari, D., Garcia, N., Lucas, E., Hamon, K., and Dhital, B. (2005). Neural Correlates of Symbolic Number Processing in Children and Adults. *Neuroreport* 16 (16), 1769–1773. doi:10.1097/01.wnr.0000183905.23396.f1

Bulthé, J., De Smedt, B., and Op de Beeck, H. P. (2014). Format-dependent Representations of Symbolic and Non-symbolic Numbers in the Human Cortex as Revealed by Multi-Voxel Pattern Analyses. *NeuroImage* 87, 311–322. doi:10.1016/j.neuroimage.2013.10.049

Bulthé, J., De Smedt, B., and Op de Beeck, H. P. (2015). Visual Number Beats Abstract Numerical Magnitude: Format-dependent Representation of Arabic Digits and Dot Patterns in Human Parietal Cortex. *J. Cogn. Neurosci.* 27 (7), 1376–1387. doi:10.1162/jocn_a_00787

Carey, S. (2004). Bootstrapping & the Origin of Concepts. *Daedalus* 133 (1), 59–68. doi:10.1162/001152604772746701

Carey, S. (2009). Where Our Number Concepts Come from. *J. Philos.* 106 (4), 220–254. doi:10.5840/JPHIL2009106418

Carey, S. (2011). Précis of the Origin of Concepts. *Behav. Brain Sci.* 34 (3), 113–124. doi:10.1017/S0140525X10000919

Clayton, S., Gilmore, C., and Inglis, M. (2015). Dot Comparison Stimuli Are Not All Alike: The Effect of Different Visual Controls on ANS Measurement. *Acta Psychol.* 161, 177–184. doi:10.1016/j.actpsy.2015.09.007

Defever, E., Sasanguie, D., Vandewaetere, M., and Reynvoet, B. (2012). What Can the Same-Different Task Tell Us about the Development of Magnitude Representations? *Acta Psychol.* 140 (1), 35–42. doi:10.1016/j.actpsy.2012.02.005

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 (2), 216–233. doi:10.1016/j.jecp.2013.04.006

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 (3), 626–641. doi:10.1037/0096-1523.16.3.626

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three Parietal Circuits for Number Processing. *Cogn. Neuropsychol.* 20 (3-6), 487–506. doi:10.1080/02643290244000239

Dehaene, S. (1996). The Organization of Brain Activations in Number Comparison: Event-Related Potentials and the Additive-Factors Method. *J. Cogn. Neurosci.* 8 (1), 47–68. doi:10.1162/jocn.1996.8.1.47

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York: Oxford University Press.

DeWind, N. K., Park, J., Woldorff, M. G., and Brannon, E. M. (2019). Numerical Encoding in Early Visual Cortex. *Cortex* 114, 76–89. doi:10.1016/j.cortex.2018.03.027

Fornaciai, M., Brannon, E. M., Woldorff, M. G., and Park, J. (2017). Numerosity Processing in Early Visual Cortex. *NeuroImage* 157, 429–438. doi:10.1016/j.neuroimage.2017.05.069

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 (1), 136–148. doi:10.1111/desc.12013

Gebuis, T., and Reynvoet, B. (2011). Generating Nonsymbolic Number Stimuli. *Behav. Res.* 43 (4), 981–986. doi:10.3758/s13428-011-0097-5

Gebuis, T., and Reynvoet, B. (2012). Continuous Visual Properties Explain Neural Responses to Nonsymbolic Number. *Psychophysiol* 49 (11), 1649–1659. doi:10.1111/j.1469-8986.2012.01461.x

Gebuis, T., Cohen Kadosh, R., and Gevers, W. (2016). Sensory-integration System rather Than Approximate Number System Underlies Numerosity Processing: A Critical Review. *Acta Psychol.* 171, 17–35. doi:10.1016/j.actpsy.2016.09.003

Gelman, R., and Gallistel, C. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

Gelman, R. (2011). The Case for Continuity. *Behav. Brain Sci.* 34 (3), 127–128. doi:10.1017/S0140525X10002712

Gentner, D., and Simms, N. (2011). Language and Analogy in Conceptual Change. *Behav. Brain Sci.* 34 (3), 128–129. doi:10.1017/S0140525X10002736

Gevers, W., Kadosh, R. C., and Gebuis, T. (2016). "Sensory Integration Theory: An Alternative to the Approximate Number System," in *Continuous Issues in Numerical Cognition*. Editor A. Henik (New York, NY: Elsevier), 405–418. doi:10.1016/B978-0-12-801637-4.00018-4

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 (6), e67374. doi:10.1371/journal.pone.0067374

Gilmore, C., Cragg, L., Hogan, G., and Inglis, M. (2016). Congruency Effects in Dot Comparison Tasks: Convex hull Is More Important Than Dot Area. *J. Cogn. Psychol.* 28 (8), 923–931. doi:10.1080/20445911.2016.1221828

Goffin, C., and Ansari, D. (2019). How Are Symbols and Nonsymbolic Numerical Magnitudes Related? Exploring Bidirectional Relationships in Early Numeracy. *Mind, Brain Educ.* 13 (3), 143–156. doi:10.1111/mbe.12206

Guillaume, M., and Van Rinsveld, A. (2018). Comparing Numerical Comparison Tasks: A Meta-Analysis of the Variability of the Weber Fraction Relative to the Generation Algorithm. *Front. Psychol.* 9, 1–9. doi:10.3389/fpsyg.2018.01694

Halberda, J., and Feigenson, L. (2008). Developmental Change in the Acuity of the "number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-Year-Olds and Adults. *Dev. Psychol.* 44 (5), 1457–1465. doi:10.1037/a0012682

Heine, A., Wißmann, J., Tamm, S., De Smedt, B., Schneider, M., Stern, E., et al. (2013). An Electrophysiological Investigation of Non-symbolic Magnitude Processing: Numerical Distance Effects in Children with and without Mathematical Learning Disabilities. *Cortex* 49 (8), 2162–2177. doi:10.1016/j.cortex.2012.11.009

Holloway, I. D., and Ansari, D. (2008). Domain-specific and Domain-General Changes in Children's Development of Number Comparison. *Dev. Sci.* 11 (5), 644–649. doi:10.1111/j.1467-7687.2008.00712.x

Holloway, I. D., and Ansari, D. (2009). Mapping Numerical Magnitudes onto Symbols: The Numerical Distance Effect and Individual Differences in Children's Mathematics Achievement. *J. Exp. Child Psychol.* 103 (1), 17–29. doi:10.1016/j.jecp.2008.04.001

Holloway, I. D., Price, G. R., and Ansari, D. (2010). Common and Segregated Neural Pathways for the Processing of Symbolic and Nonsymbolic Numerical Magnitude: An fMRI Study. *Neuroimage* 49 (1), 1006–1017. doi:10.1016/j.neuroimage.2009.07.071

Hyde, D. C., and Spelke, E. S. (2009). All Numbers Are Not Equal: An Electrophysiological Investigation of Small and Large Number Representations. *J. Cogn. Neurosci.* 21 (6), 1039–1053. doi:10.1162/jocn.2009.21090

Hyde, D. C., and Spelke, E. S. (2011). Neural Signatures of Number Processing in Human Infants: Evidence for Two Core Systems Underlying Numerical Cognition. *Dev. Sci.* 14 (2), 360–371. doi:10.1111/j.1467-7687.2010.00987.x

Hyde, D. C., and Spelke, E. S. (2012). Spatiotemporal Dynamics of Processing Nonsymbolic Number: An Event-related Potential Source Localization Study. *Hum. Brain Mapp.* 33 (9), 2189–2203. doi:10.1002/hbm.21352

Kadosh, R. C., Lammertyn, J., and Izard, V. (2008). Are Numbers Special? an Overview of Chronometric, Neuroimaging, Developmental and Comparative Studies of Magnitude Representation. *Prog. Neurobiol.* 84 (2), 132–147. doi:10.1016/j.pneurobio.2007.11.001

Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early Numerical Development and the Role of Non-symbolic and Symbolic Skills. *Learn. Instruction* 25, 95–103. doi:10.1016/j.learninstruc.2012.12.001

Krajcsi, A., Lengyel, G., and Kojouharova, P. (2016). The Source of the Symbolic Numerical Distance and Size Effects. *Front. Psychol.* 7, 1–16. doi:10.3389/fpsyg.2016.01795

Krajcsi, A., Lengyel, G., and Kojouharova, P. (2018). Symbolic Number Comparison is Not Processed by The Analog Number System: Different Symbolic and Non-Symbolic Numerical Distance and Size Effects. *Front. Psychol.* 9, 124. doi:10.3389/fpsyg.2018.00124

Landgraf, S., Meer, E., and Krueger, F. (2010). Cognitive Resource Allocation for Neural Activity Underlying Mathematical Cognition: A Multi-Method Study. *ZDM Math. Educ.* 42, 579–590. doi:10.1007/s11858-010-0264-7

Landy, D., Allen, C., and Anderson, M. L. (2011). Conceptual Discontinuity Involves Recycling Old Processes in New Domains. *Behav. Brain Sci.* 34 (3), 136–137. doi:10.1017/S0140525X10002153

Leibovich, T., and Ansari, D. (2016). The Symbol-Grounding Problem in Numerical Cognition: A Review of Theory, Evidence, and Outstanding Questions. *Can J. Exp. Psychol.* 70 (1), 12. doi:10.1037/cep0000070

Leibovich, T., and Henik, A. (2014). Comparing Performance in Discrete and Continuous Comparison Tasks. *Q. J. Exp. Psychol.* 67 (5), 899–917. doi:10.1080/17470218.2013.837940

Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From "Sense of Number" to "Sense of Magnitude": The Role of Continuous Magnitudes in Numerical Cognition. *Behav. Brain Sci.* 40, 1–62. doi:10.1017/s0140525x16000960

Libertus, M. E., Woldorff, M. G., and Brannon, E. M. (2007). Electrophysiological Evidence for Notation independence in Numerical Processing. *Behav. Brain Funct.* 3 (1), 1–15. doi:10.1186/1744-9081-3-1

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 (4), 635–641. doi:10.1037/a0027248

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 (3), 440–457. doi:10.1037/dev0000445

Marinova, M., Sasanguie, D., and Reynvoet, B. (2018). Symbolic Estrangement or Symbolic Integration of Numerals with Quantities: Methodological Pitfalls and a Possible Solution. *PLoS One* 13 (7), e0200808. doi:10.1371/journal.pone.0200808

Marinova, M., Sasanguie, D., and Reynvoet, B. (2021). Numerals Do Not Need Numerosities: Robust Evidence for Distinct Numerical Representations for Symbolic and Non-symbolic Numbers. *Psychol. Res.* 85, 764–776. doi:10.1007/s00426-019-01286-z

The Mathworks, Inc (2017). *Matlab R2017a*. Massachusetts: Natick.

Nys, J., and Content, A. (2012). Judgement of Discrete and Continuous Quantity in Adults: Number Counts!. *Q. J. Exp. Psychol.* 65 (4), 675–690. doi:10.1080/17470218.2011.619661

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intelligence Neurosci.* 2011, 1–9. doi:10.1155/2011/156869

Park, J., DeWind, N. K., Woldorff, M. G., and Brannon, E. M. (2016). Rapid and Direct Encoding of Numerosity in the Visual Stream. *Cereb. Cortex* 26 (2), bhv017–763. doi:10.1093/cercor/bhv017

Park, J., DeWind, N. K., and Brannon, E. M. (2017). Direct and Rapid Encoding of Numerosity in the Visual Stream. *Behav. Brain Sci.* 40, 2532–2544. doi:10.1017/S0140525X16002235

Park, J. (2018). A Neural Basis for the Visual Sense of Number and its Development: A Steady-State Visual Evoked Potential Study in Children and Adults. *Dev. Cogn. Neurosci.* 30, 333–343. doi:10.1016/j.dcn.2017.02.011

Paulsen, D. J., and Neville, H. J. (2008). The Processing of Non-symbolic Numerical Magnitudes as Indexed by ERPs. *Neuropsychologia* 46 (10), 2532–2544. doi:10.1016/j.neuropsychologia.2008.04.003

Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron* 53 (2), 293–305. doi:10.1016/j.neuron.2006.11.022

Piazza, M., De Feo, V., Panzeri, S., and Dehaene, S. (2018). Learning to Focus on Number. *Cognition* 181, 35–45. doi:10.1016/j.cognition.2018.07.011

Pinel, P., Dehaene, S., Rivière, D., and LeBihan, D. (2001). Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task. *NeuroImage* 14 (5), 1013–1026. doi:10.1006/nimg.2001.0913

Reynvoet, B., and Sasanguie, D. (2016). The Symbol Grounding Problem Revisited: A Thorough Evaluation of the ANS Mapping Account and the Proposal of an Alternative Account Based on Symbol-Symbol Associations. *Front. Psychol.* 07, 1–11. doi:10.3389/fpsyg.2016.01581

Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012). Association between Basic Numerical Abilities and Mathematics Achievement. *Br. J. Dev. Psychol.* 30 (2), 344–357. doi:10.1111/j.2044-835X.2011.02048.x

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 (3), 418–431. doi:10.1016/j.jecp.2012.10.012

Sasanguie, D., De Smedt, B., and Reynvoet, B. (2017). Evidence for Distinct Magnitude Systems for Symbolic and Non-symbolic Number. *Psychol. Res.* 81 (1), 231–242. doi:10.1007/s00426-015-0734-1

Sassenhagen, J., and Draschkow, D. (2019). Cluster-based Permutation Tests of MEG/EEG Data Do Not Establish Significance of Effect Latency or Location. *Psychophysiology* 56 (6), e13335. doi:10.1111/psyp.13335

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of Non-symbolic and Symbolic Numerical Magnitude Processing with Mathematical Competence: a Meta-Analysis. *Dev. Sci.* 20 (3), e12372–16. doi:10.1111/desc.12372

Sokolowski, H. M., Fias, W., Mousa, A., and Ansari, D. (2017). Common and Distinct Brain Regions in Both Parietal and Frontal Cortex Support Symbolic and Nonsymbolic Number Processing in Humans: A Functional Neuroimaging Meta-Analysis. *Neuroimage* 146, 376–394. doi:10.1016/j.neuroimage.2016.10.028

Soltész, F., and Szűcs, D. (2014). Neural Adaptation to Non-symbolic Number and Visual Shape: an Electrophysiological Study. *Biol. Psychol.* 103, 203–211. doi:10.1016/j.biopsycho.2014.09.006

Soltész, F., Szűcs, D., Dékány, J., Márkus, A., and Csépe, V. (2007). A Combined Event-Related Potential and Neuropsychological Investigation of Developmental Dyscalculia. *Neurosci. Lett.* 417, 181–186. doi:10.1016/j.neulet.2007.02.067

Spelke, E. S. (2011). Quinian Bootstrapping or Fodorian Combination? Core and Constructed Knowledge of Number. *Behav. Brain Sci.* 34 (3), 149–150. doi:10.1017/S0140525X10002220

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

Temple, E., and Posner, M. I. (1998). Brain Mechanisms of Quantity Are Similar in 5-Year-Old Children and Adults. *Proc. Natl. Acad. Sci.* 95 (13), 7836–7841. doi:10.1073/pnas.95.13.7836

Van Hoogmoed, A. H., and Kroesbergen, E. H. (2018). On the Difference between Numerosity Processing and Number Processing. *Front. Psychol.* 9, 1–13. doi:10.3389/fpsyg.2018.01650

Van Opstal, F., and Verguts, T. (2011). The Origins of the Numerical Distance Effect: The Same-Different Task. *J. Cogn. Psychol.* 23 (1), 112–120. doi:10.1080/20445911.2011.466796

Van Opstal, F., Gevers, W., De Moor, W., and Verguts, T. (2008). Dissecting the Symbolic Distance Effect: Comparison and Priming Effects in Numerical and Nonnumerical Orders. *Psychon. Bull. Rev.* 15 (2), 419–425. doi:10.3758/PBR.15.2.419

Van Rinsveld, A., Guillaume, M., Kohler, P. J., Schiltz, C., Gevers, W., and Content, A. (2020). The Neural Signature of Numerosity by Separating Numerical and Continuous Magnitude Extraction in Visual Cortex with Frequency-Tagged EEG. *Proc. Natl. Acad. Sci. USA* 117 (11), 5726–5732. doi:10.1073/pnas.1917849117

Verguts, T., and Van Opstal, F. (2005). Dissociation of the Distance Effect and Size Effect in One-Digit Numbers. *Psychon. Bull. Rev.* 12 (5), 925–930. doi:10.3758/BF03196787

Xu, F., and Arriaga, R. I. (2007). Number Discrimination in 10-Month-Old Infants. *Br. J. Dev. Psychol.* 25, 103–108. doi:10.1348/026151005X90704

Xu, F., and Spelke, E. S. (2000). Large Number Discrimination in 6-Month-Old Infants. *Cognition* 74 (1), 1–11. doi:10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., and Goddard, S. (2005). Number Sense in Human Infants. *Developmental Sci* 8, 88–101. doi:10.1111/j.1467-7687.2005.00395.x

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.

ORIGINAL RESEARCH
published: 19 July 2021
doi: 10.3389/feduc.2021.683672

Effects of Gender on Basic Numerical and Arithmetic Skills: Pilot Data From Third to Ninth Grade for a Large-Scale Online Dyscalculia Screener

Pekka Räsänen[1,2]*, Pirjo Aunio[3], Anu Laine[3], Airi Hakkarainen[4], Eija Väisänen[3], Jonatan Finell[5], Teemu Rajala[1], Mikko-Jussi Laakso[1] and Johan Korhonen[6]

[1]Faculty of Technology, University of Turku, Turku, Finland, [2]Division of Social and Health Services, Helsinki, Finland, [3]Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [4]Open University, University of Helsinki, Helsinki, Finland, [5]Department of Applied Educational Science, Umeå University, Umeå, Sweden, [6]Faculty of Education and Welfare Studies, Åbo Akademi University, Turku, Finland

OPEN ACCESS

Edited by:
Korbinian Moeller,
Loughborough University,
United Kingdom

Reviewed by:
Alberto Crescentini,
University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Switzerland
Anies Al-Hroub,
American University of Beirut, Lebanon

***Correspondence:**
Pekka Räsänen
pjrasa@utu.fi

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

Received: 21 March 2021
Accepted: 28 May 2021
Published: 19 July 2021

Citation:
Räsänen P, Aunio P, Laine A, Hakkarainen A, Väisänen E, Finell J, Rajala T, Laakso M-J and Korhonen J (2021) Effects of Gender on Basic Numerical and Arithmetic Skills: Pilot Data From Third to Ninth Grade for a Large-Scale Online Dyscalculia Screener. Front. Educ. 6:683672. doi: 10.3389/feduc.2021.683672

In this study, we analyzed the development and effects of gender on basic number skills from third to ninth grade in Finland. Because the international comparison studies have shown slightly different developmental trends in mathematical attainment for different language groups in Finland, we added the language of education as a variable in our analysis. Participants were 4,265 students from third to ninth grade in Finland, representing students in two national languages (Finnish, n = 2,833, and Swedish, n = 1,432). Confirmatory factor analyses showed that the subtasks in the dyscalculia screener formed two separate factors, namely, number-processing skills and arithmetic fluency. We found a linear development trend across age cohorts in both the factors. Reliability and validity evidence of the measures supported the use of these tasks in the whole age group from 9 to15 years. In this sample, there was an increasing gender difference in favor of girls and Swedish-speaking students by grade levels in number-processing skills. At the same time, boys showed a better performance and a larger variance in tasks measuring arithmetic fluency. The results indicate that the gender ratio within the group with mathematical learning disabilities depends directly on tasks used to measure their basic number skills.

Keywords: learning disabilities, number sense, arithmetic fluency, language, gender differences, mathematics, variance ratio, basic number skills

INTRODUCTION

The easy access to the internet and computer technology is changing the way we assess mathematical learning disabilities (MLD). There is a long history of using computerized tasks to assess numerical skills in research. In addition, many international and national assessments, such as OECD PISA studies, are nowadays conducted online. However, a transformation of this research into practical diagnostic tools for clinical educational psychology is still in its infancy (Conole and Warburton, 2005; Räsänen et al., 2015; Molnár andCsapó, 2019; Räsänen et al., 2019).

It has been shown repeatedly that basic number skills form the foundations for learning more complex mathematical skills (Butterworth, 2005; Jordan et al., 2009; Li et al., 2018), and early numerical skills predict later achievement in mathematics (Zhang et al., 2017; Blume et al., 2021).

Furthermore, research has shown that weak basic numerical skills form the core deficit in MLD in groups of younger and older students (De Smedt et al., 2013; Zhang et al., 2017). Therefore, assessment of basic numerical skills should be part of every clinical evaluation of MLD.

However, there is not much information on how the basic number skills develop during the school years. Halberda et al. (2012) showed that the fastest development phase in ANS (approximate number system) is between 11 to 16 years of age, not at a younger age range, as expected from such a fundamental skill. ANS, which is typically measured with nonsymbolic number comparison tasks, has a small but significant correlation with mathematical skills in all age groups. However, this skill does not seem to be a reliable task to differentiate children with and without MLD in groups under ten years of age (De Smedt et al., 2013). Brankaer et al. (2017) found that students' symbolic number comparison skills improved during the whole primary school (grades 1–6) and were consistently related to students' math performance in all grades.

There are no commonly agreed models of how the basic number skills should be defined or categorized and what tasks should be implemented into a clinical test battery. Aunio and Räsänen (2016) suggested that the core set of skills that should be measured could be clustered into four groups: number sense, counting skills, arithmetic, and understanding mathematical relations. This division did get some support from a factor analytic study with one test battery (Hellstrand et al., 2020). However, this model has not been replicated with other test batteries. Reigosa-Crespo et al. (2012), who screened MLD from over eleven thousand children from second to ninth grade with a computerized test battery, divided their tasks into two groups, namely, basic numerical skills (enumeration and number comparison) and arithmetic fluency. However, the division of tasks into these categories was not based on data analysis. What is clear is that even the basic numerical processing is made up of many different components with different developmental trajectories and relationships to arithmetic achievement (Lyons et al., 2014).

Gender

The international comparison studies on mathematical attainment have shown significant differences in mathematical performances between countries, educational cultures, types of schools, socioeconomic groups, and genders (e.g., OECD, 2019; Mullis et al., 2020). To look at the gender differences in mathematical skills, Reilly et al. (2017) analyzed the results of 45 countries from the 2011 Trends in Mathematics and Science Study (TIMSS). They found small- to medium-sized gender differences for most individual nations with a substantial variation (d =−0.60 to +0.31). The direction varies, and there seem to be no global gender differences, but gender differences seem to be immutable. These international comparison studies of attainment focus on a variety of mathematical skills, mainly concentrating on curriculum-based contents of more complex mathematics and its different applications learned following the curricular plans of the local school systems. Therefore, it is not surprising that there are significant differences between educational cultures, socioeconomic groups, and genders in mathematical skills. However, the differences in more complex skills do not directly tell us if there are differences in basic numerical skills. Surprisingly, only a few studies on the effects of cultural factors, such as language, and only slightly more about gender effects on basic numerical skills have been published.

Stereotypes that girls lack mathematical ability persist and are widely held by parents and teachers (Hyde et al., 2008). Many studies aim to find explanations for this "male advantage." Typically, in addition to gender stereotypes, explanations for early-grade gender differences have been searched from domain-general and domain-specific cognitive variables. For example, van Tetering et al. (2019) showed that boys outperformed girls in mathematics in most grade levels within children from 7 to 12 years old. At the same time, boys also showed a better performance in spatial mental rotation skills. The authors concluded that their results "suggest that interventions that stimulate the development of spatial skills may facilitate mathematical achievements, especially of young girls" (see also Rosselli et al., 2009). Similarly, Royer et al. (1999) showed in a series of analyses that arithmetic, favoring boys, could explain the gender differences in more complex math performance.

If, for example, boys would outperform girls also in basic number skills, this would lend support to the stereotype that boys have an early cognitive advantage (such as spatial skills or arithmetic fluency) that would explain the differences in more complex mathematical skills later on. However, if there would not be differences between girls and boys on basic number skills, it would suggest that both genders are equally equipped to acquire more complex math skills (Bakker et al., 2019; Hutchison et al., 2019). The reversed results favoring girls might reflect a cognitive advantage supporting girls. For example, Wei et al. (2012) found in a study with 8- to 11-year-old Chinese children that verbal fluency explained the girls' better arithmetic skills. The gender differences might also mean that the fundamental number skills are strongly malleable to cultural effects. The relationship between basic number skills and more complex mathematical skills may be more reciprocal than expected. The gender differences in basic number skills could also reflect how mathematical skills develop in general within each educational culture.

There is an extensive number of studies on gender differences in school-related mathematical skills. Anastasi (1958) showed that boys outperform girls in mathematics during the elementary school years with some exceptions. For example, girls excelled in computational fluency, while boys performed better on more cognitively demanding tasks such as problem-solving. The early research reviews reported consistent gender differences in mathematical achievement (Fennema, 1974; Halpern, 1986). In the 1990s, Hyde et al., 1990 showed in their extensive meta-analysis of 100 studies (over 3 million subjects) that the gender gap in mathematical achievement had diminished over time, and the recent studies have shown that in developed countries, the genders show an equal aptitude for mathematics (Hyde et al., 2008; Lindberg et al., 2010).

Recently in some OECD countries, there has been a trend that females have started to outperform males at most levels of education and are better represented in universities (OECD, 2015). In some countries, such as Finland, where we conducted this study, girls have also started to outperform boys in school mathematics at the upper grades. However, the gender gaps favoring males have persisted, for example, in average income, employment in prestigious occupations, and leadership roles (CEDA, 2013; Goldin, 2014). Likewise, even though the gender gap in educational achievements would have narrowed or even reversed, the differences, in favor of men, have remained in self-concept and self-promotion (Parker et al., 2018). These last-mentioned noncognitive factors may affect the career choices to STEM disciplines (O'Dea et al., 2018).

Cross-cultural studies have shown that even though there would be differences in school mathematics, there would be no systematic gender differences in basic numerical or calculation skills in younger age groups (Geary et al., 1996; Aunio et al., 2006). Geary, with his colleagues, tested children from kindergarten through third grade from China and the United States using single-digit addition and found no gender effects on the accuracy of performance in either country. Shen et al. (2016) compared arithmetic skills of 7 year-olds in three countries finding that the gender differences varied from one country to another. In simple arithmetic tasks, the gender differences were visible in the strategies but not in the accuracy. In more complex tasks, the gender effect varied by country, reflecting that the educational context may play a role in gender differences in mathematics (Shen et al., 2016).

Hutchison et al. (2019) were the first to publish a systematic large-scale study on gender differences at school-age in tasks measuring basic number skills. They studied 6- to 13-year-old children (grades 1–6) with a large battery of tasks in seven different primary schools in Netherlands. The tasks to measure the basic number skills were similar to those typically used in studies aiming to grasp the fundamental features of MLD (Bartelet et al., 2014; Lyons et al., 2014). They summarized their results to "provide strong evidence of gender similarities on the majority of basic numerical tasks measured, suggesting that a male advantage in foundational numerical skills is the exception rather than the rule." Moreover, they concluded that this is strong support for the idea that boys and girls are equally equipped with basic numerical competencies and should be equally capable of acquiring complex mathematical skills. Kersey et al. (2018) came to the same conclusion in their large-scale analysis of gender differences. They used different datasets of basic numerical skills collected in different studies of children from 6 months to 8 year-olds.

The older studies that reported gender differences in tasks measuring basic number skills had very mixed results. Krinzinger et al. (2012) studied children at primary school and found that there was a gender difference favoring boys on single- and especially on multi-digit number comparison, while another study (Wei et al., 2012) found an opposite result with a similar task and an eight times larger sample (N = 1,156). Rosselli et al. (2009) did not find any gender differences in their analysis on a number comparison, reading numbers, writing numbers, and ordering numbers in a sample of 526 7–16 year-olds.

Like mentioned earlier, the male advantage in mathematical skills has often been connected to spatial skills (van Tetering et al., 2019). There is strong evidence of male advantage in some aspects of spatial cognition (Halpern et al., 2007; Levine et al., 2016). Spatial skills have been shown to explain mathematical skills (Resnick et al., 2019), as well as the development of mathematical skills (Zhang et al., 2017). Therefore, it is not a surprise to find male advantage in numerical tasks that are based on spatial representations of numbers, such as the SNARC effect (Spatial Numerical Association of Response Codes) and number line estimation tasks. Boys seem to show a larger SNARC effect (Bull et al., 2013), and their estimations are more accurate in a number line estimation task (Thompson and Opfer, 2008; Gunderson et al., 2012; Bull et al., 2013; Reinert et al., 2016). When moving away from spatial numerical tasks to symbolic tasks, the picture of gender differences or similarities becomes blurry. Bull, Cleland, and Mitchell (2013) studied an adult sample. They found that males were faster in discriminating between two numbers and that only females displayed a numerical distance effect (logarithmic vs. linear representation). They suggested that males would have a more accurate representation of number/magnitude, which helps them discriminate between numbers closer to each other. However, studies with children have not replicated this finding with similar types of number comparison tasks (Wei et al., 2012; Krinzinger et al., 2012; Lyons et al., 2015). One factor that may explain the differences between studies is the large variety in the tasks used to assess the gender differences. Another confounding factor is the difference in the age groups of the studies. The studies that showed conflicting results focused mainly on children between the ages of 6–10 years.

Hutchison et al. (2019) analyzed gender differences in a study with children from first to sixth grade (7–13 year-olds, N = 1,463). Their test battery consisted of two numerical comparison tasks (symbolic and nonsymbolic), two matching tasks (visual and auditory), number line estimation, numerical ordering, counting, and two arithmetic tasks (addition/subtraction, multiplication/division). The only systematic gender effect found was in a number line estimation task, where the effect was strong at the early grades but disappeared at the sixth grade. The gender similarity was a systematic finding in their study.

Thus far, Reigosa-Crespo et al. (2012) had the most extensive sample to measure basic numerical and arithmetic skills. They screened over eleven thousand children from second to ninth grade. Unfortunately, they did not report the gender differences directly, but only the ratios within the low- and high-performing groups in the whole sample. They divided their tasks into two subskills: basic numerical skills (enumeration and number comparison) and arithmetic fluency. They found a higher prevalence of boys than girls at the lower end of efficiency in the basic numerical skills. Boys were two times more likely to have a deficit in basic numerical skills compared to girls. In addition, there were four times more boys than girls in the group, which had a deficit both in basic numerical skills and arithmetical fluency. They did not find differences between genders at the higher end of efficiency in enumeration or number comparison tasks but failed to report the results on their arithmetic tasks.

Variance Ratio

While most of the studies on basic number skills find only small or nonexisting differences in means between the genders, in the context of assessing MLD, the differences in variance may be more critical because the differences in variance affect the ends of the skill distribution. There is a long history of analyzing gender differences in variance of cognitive and academic skills (Maccoby andJacklin, 1974; Feingold, 1992). Feingold (1992) summarized that males were more variable than females in quantitative ability and spatial visualization, while there were no differences in variance in verbal tests, short-term memory, abstract reasoning, and perceptual speed. Nowell and Hedges (1998), in their extensive analysis on the datasets of mathematical attainment in the US national assessment, showed that the variance ratio (VR, male variance/female variance) had not changed in mathematics from 1978 to 1994, constantly showing a larger variance for males (1.05–1.42 in mathematics in their report). While the gender gap in means seems to be closing, the variance ratio has been more stable. In the latest studies, while the gender difference in means is no longer significant, the larger male variance in mathematical skills is still found. However, the difference is not so extent that it could alone explain the overrepresentation of males in the STEM field (O'Dea et al., 2018). Interestingly, Penner and Paret (2008) showed that differences in the variance exist already at preschool age.

Therefore, even though the gender similarity hypothesis in basic numerical skills (Bakker et al., 2019; Hutchison et al., 2019) would be systematically replicated and confirmed, the differences in variance could still produce significant gender differences within the extremes. However, until now, also these results have been very mixed at the lower end of the distribution. While Reigosa-Crespo et al. (2012) reported up to four times more boys having MLD than girls, only some studies have agreed on this (Badian, 1983; Ramaa andGowramma, 2002; Barbaresi et al., 2005). Some studies have shown an equal number of genders in the group of MLD (Lewis et al., 1994; Mazzocco and Myers, 2003; Koumoula et al., 2004; Devine et al., 2013), while some studies have shown the reverse gender difference, i.e., a larger number of girls than boys with MLD (Shalev et al., 2000; Dirks et al., 2008). There is a need to look at this question at a task level whether the differences in variances are systematically similar from one task to another.

Culture and Language

Our dataset was collected in Finland. The closest comparison to our dataset where similar measures were used is Hutchison's (2019) study conducted in Netherlands. Finland and Netherlands have both been high-performing countries in international mathematical comparison studies. There have not been significant differences in how girls and boys perform in school mathematics in these countries. For example, in the latest TIMSS study of 14–15 year-olds, in Netherlands, boys were a nonsignificant +12 points better than girls. At the same time, in Finland, there were no significant gender differences in mathematics at this age, but girls' average was slightly above those of the boys.

Finland has two official languages, Finnish and Swedish. The Finnish-speaking schools used to perform slightly better than the Swedish-speaking schools (Kupari et al., 2012). Even though in Finland, there are no significant socioeconomic or educational differences between the schools in the system of free public education (Kupiainen et al., 2009). All children in Finland participate in the same public education offered by similarly university-trained teachers, and they all follow exactly the same national curriculum framework.

There has been a shift in mathematical attainment between the genders and between the two language groups in Finland during the last 2 decades. Today, girls perform better than boys, and the Swedish-speaking minority performs better than the Finnish-speaking majority. In the latest TIMSS 2018 study, in the fourth-grade sample, there were no differences in mathematics between the language groups nor genders (Vettenranta et al., 2020a). However, in the eighth grade, the Swedish-speaking sample was slightly better in mathematics, especially the Swedish-speaking girls (Vettenranta et al., 2020b). The trends of improvement of girls' performance levels compared to that of boys' and the improvement of the Swedish-speaking minority compared to the Finnish-speaking majority are also visible in the PISA data (**Figure 1**). In PISA data, the main effect producing these trends in Finland has been that Swedish-speaking girls are the only group that has not shown a similar constant decline in their math performance as the other groups (OECD 2015; OECD 2019).

Summary

In this study, we analyzed the effects of gender in basic number skills from third to ninth grade to add one educational culture, Finland, to the small number of studies looking at the gender differences in basic numerical skills. Because the international comparison studies have shown slightly different developmental trends in mathematical attainment for different language groups in Finland, we added the education language in the school as a variable into our analysis. There are two main reasons, one theoretical and one practical, why we are interested in the gender differences and the effects of the language group when we assess the basic number skills. First, the previous studies have shown very mixed results indicating no systematic differences between the genders. The most systematic study until now indicates that gender similarity is the rule and the differences an exception (Hutchison et al., 2019). However, another possible explanation for the mixed results is that there could be a reciprocal relationship between basic number skills and school mathematics. The gender differences and gender similarities in basic number skills may reflect the results of the curriculum-based assessments. Therefore, we would find increasing gender differences favoring Swedish-speaking girls in the older age groups, as has been the trend in the international and national achievement studies.

Second, a practical reason for this analysis is that our data collecting was part of a process to develop an online test battery for clinical use. This study is our first pilot to test both the online technology in practice and investigate the suitability of the tasks for the test battery for screening mathematical learning

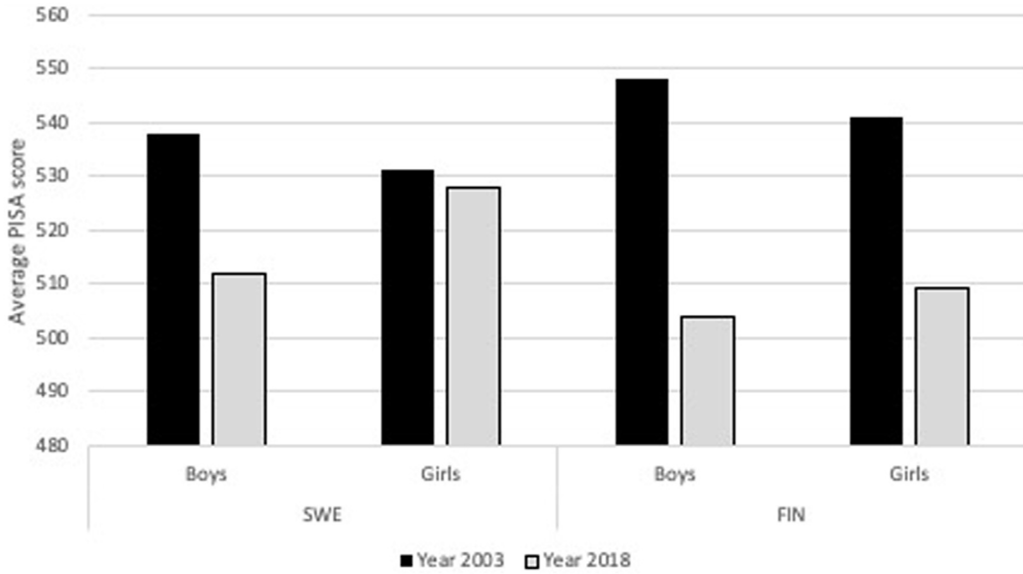


FIGURE 1 | The average mathematics attainment in PISA studies of the Finnish and Swedish-speaking samples in Finland in 2003 and 2018 (Source: OECD).

difficulties. Systematic and significant differences due to gender or language would mean that we should take these differences into account in the forthcoming standardization process of our test. Any differences in means or variation would affect the gender ratio of those diagnosed as having MLD. Substantial differences in some tasks would require us to consider providing different norms for different subgroups. From the clinical perspective, the gender differences in the extremes are even more important than the differences in means. Therefore, we also report here the gender ratios in the extreme values. The previous studies on MLD have shown all three possibilities in the gender ratios. More information is needed to see how the different tasks affect the ratio of males vs. females in the extremes. Therefore, our results will also function as information for others who aim to develop standardized test batteries for screening MLD.

To reliably compare different groups in basic numerical skills, we first need to ensure that our measure 1) shows adequate reliability, 2) structural validity, and 3) measurement invariance across groups (Finnish vs. Swedish; boys vs. girls). Hence, the analyses start with establishing reliability and validity evidence of our test battery. Second, we will look at the trends of the gender differences at different grade levels controlling for the language of instruction. Last, we will look at the variance and the gender ratio in the low- and high-performing extremes.

MATERIALS AND METHODS

This study is part of a larger FUNA (Functional Numeracy Assessment) project to develop a test battery to assess basic numerical and mathematical skills (see http://oppimisanalytiikka.fi/funa). These data are from a subproject to develop a screening test battery for mathematical learning disabilities (dyscalculia). When ready, the FUNA dyscalculia battery (FUNA-DB) will consist of seven tasks measuring basic number processing and arithmetic skills. The test battery runs on an online educational platform offered to schools in Finland by the Center of Learning Analytics at the University of Turku. The system can offer the contents on an internet browser and collect all user interactions and their timings for further analysis. The system works on all operating systems and machines (computers, tablets, and mobile devices) (more information about the platform in English, see http://eduten.com).

Participants

We collected the data for this pilot study with the help of voluntary teachers and schools. Three methods were used to find volunteers: First, we held three two-day teacher training on dyscalculia, one in North, one in Central, and one in South Finland. The aim of the teacher training offered was to encourage teachers to participate in the data collecting. The teacher training consisted of two days of lectures about dyscalculia (neuropsychology and intervention methods, instructions on how to conduct the assessment and how to interpret the test results), and an assessment of classes of pupils at the schools of the participating teachers using FUNA-DB. Second, we searched for additional voluntary teachers via an advertisement in a newsletter that reaches almost all schools in Finland. Third, we took direct contacts to schools to add the number of schools to the Swedish-speaking sample.

The pupils participated in the study anonymously. The teacher informed the number of girls and boys, their grade levels, and the language of the school to our research assistant. The assistant generated an equal number of random logins/passwords that contained a hidden code for gender, grade, and language. The teacher gave these codes to the children based on their gender and grade levels. These three variables were the only pieces of

TABLE 1 | The sample sizes in different cells of grade, gender, and language.

Grade	Finnish		Swedish		Total		
	Boys	Girls	Boys	Girls	Boys	Girls	Total
3	198	191	129	148	327	339	666
4	137	154	163	185	300	339	639
5	178	184	117	91	295	275	570
6	191	189	134	129	325	318	643
7	282	240	93	87	375	327	702
8	273	269	53	58	326	327	653
9	183	164	19	26	202	190	392
Total	1,442	1,391	708	724	2,150	2,115	4,265

information that were obtained from the children. Each teacher received feedback from the performance of each of their pupil who participated in the study. The teacher received the stanine scores based on the results of the total sample at each grade level. No other feedback or rewards were given.

The study was conducted as a collaboration with schools from tens of municipalities. Research permission and ethical approval were applied from the local educational research committee of each municipality separately. A research permit was obtained, and the participating pupils' parents were informed about the study following the instructions and policy of each municipal school authority.

The total sample size was 4,265 pupils from third to ninth grade in two national languages (Finnish, $n = 2{,}833$, and Swedish, $n = 1{,}432$) in Finland. In **Table 1**, there is a summary of the number of pupils broken by grade, gender, and language.

The Tasks and the Assessments

In this pilot study, data were collected using seven tasks. However, due to an experimenter error, only six of those tasks are used in this analysis, namely, Number comparison, Digit dot matching, Number series, Single-digit addition, Single-digit subtraction, and Multi-digit calculations (addition/subtraction). In the Number series and Multi-digit Calculation tasks, there were five different parallel versions of the task, which were randomly allocated to the subjects. It means that in the same classrooms, the subjects did slightly different versions of the test batteries.

The teachers were given word-by-word instructions on how to conduct the assessments. After login the pupils were able to proceed in their own speed with the tasks without further instructions from the teacher or other interruptions. Each task started with instructions and had a practice task with 4–5 practice items before that actual task.

The median reaction times and accuracy by grade, gender, and language are presented in the supplementary materials.

Number Comparison

Two single-digit two Arabic numbers were presented on the screen, and the subject was asked to press as soon as possible the button (or key if using a computer) on the same side where the larger of the two numbers was. Each subject was shown a total of 52 items, of which ten were removed from the score calculation (items containing either 1 or 9). The remaining 42 items consisted of pairs of numbers from two to eight. The presentation order of the number pairs for each subject was fully randomized. The score used in the analysis was an efficiency score (the median reaction time of the correct responses divided by the percentage of correct responses). Split-half reliability of the task was Spearman-Brown = 0.924, Guttman split-half = 0.845.

Digit Dot Matching Task

In this equivalence task, the subjects were asked to press as fast as they could one of the two buttons ("same" or "different" or one of the two keys if using a computer) based on the equivalence of the quantities presented in the stimuli. There was an Arabic number on the left side and a randomly organized dot pattern on the right side. The matching pairs (all numbers from 1 to 9) were presented twice, and the remaining nonmatching items were divided into small-difference (e.g., 3 vs. 4) and large-difference items (e.g., 3 vs. 8). A total of 42 items were presented. The score used in the analysis was an efficiency score (the median reaction time of the correct responses divided by the percentage of correct responses). Split-half reliability of the task was Spearman-Brown = 0.756, Guttman split-half = 0.756.

Number Series

A total of 20 series of numbers were presented in order of difficulty. In each item, there were four numbers, and the subject was asked to continue the series based on the rule that the four numbers formed. There were five parallel versions of the series, each containing five same anchor items. The maximum time to solve the problems was 5 min. The score used in the analysis was an efficiency score (the median reaction time of the correct responses divided by the percentage of correct responses). Split-half reliability of the task was Spearman-Brown = 0.803, Guttman split-half = 0.707.

Single-Digit Addition

All 81 single-digit number combinations from 1 to 9 were presented to the subject as an addition (e.g., 3 + 4 = _) in a quasi-random order. There was a digital number pad on the screen which the subject could use to type in the answer (also, the number keys on a computer keyboard could be used). The subjects were instructed to answer as many items as they could during the 2 min time limit. During the last 15 s of the task, there appeared a warning about the ending of the response time. The score was the number of correct items in 2 min. Split-half reliability of the task was Spearman-Brown = 0.995, Guttman split-half = 0.995.

Single-Digit Subtraction

The reverse of the single-digit addition task was presented as subtractions (e.g., 7–3 = _; the answer of the addition task as the minuend). All 81 number combinations were presented in a quasi-random order to the subject. There was a digital number pad on the screen which the subject could use to type in the answer (also, the number keys on a computer keyboard could be used). The subjects were instructed to answer as many items as they could during the 2 min time limit. During the last 15 s of the task, there appeared a warning about the ending of the response time. The score was the number of correct items in 2 min. Split-half reliability of the task was Spearman-Brown = 0.993, Guttman split-half = 0.993.

Multi-Digit Addition and Subtraction

Five different series of addition and subtraction tasks were created from two-to four-digit numbers (e.g., 20 + 50 = _, 320–80 = _) in order of difficulty (i.e., the number of steps required to calculate the answer). Each item in the parallel versions was created to have a matching pair in the other series. Twenty out of the 80 items were anchor items across the series. The subjects were instructed to answer as fast as possible. The score was the number of correct items in 3 min. During the last 15 s of the task, there appeared a warning about the ending of the response time. Split-half reliability of the task was Spearman-Brown = 0.993, Guttman split-half = 0.992.

Statistical Analysis

The analyses were conducted with the SPSS (version 26) and Mplus (version 8.4) statistical software. The factor structure of the FUNA-DB was explored utilizing confirmatory factor analysis (CFA). More specifically, a one-factor model that assumes that all tasks load on an overall basic numerical skills factor was compared to a two-factor model consisting of a number processing factor (number comparison, digit-dot matching) and an arithmetic fluency factor (Number Series, Single-digit Addition, Single-digit Subtraction, Multi-digit Calculations). Measurement invariance was tested with multigroup CFA. In multigroup CFA, a series of nested models are fitted to the data where the endpoints are the least restrictive model with no invariance constraints and the most restrictive model where all parameters are forced to equality across groups (Bollen, 1989). In all analyses, we used the Full information maximum likelihood (FIML) that uses all available data as the estimator. We used chi-square (X^2), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA) as model-fit indicators. The CFI and TLI vary along a 0-to-1 continuum, and values greater than 0.90 and 0.95 typically reflect acceptable and excellent fit to the data, respectively. RMSEA values of less than 0.05 and 0.08 reflect a close fit and a reasonable fit to the data, respectively (Marsh, Hau, and Wen, 2004). To compare nested models, we looked at the change in CFI and RMSEA (Chen, 2007). According to Chen (2007), support for the more parsimonious model requires a change in CFI (ΔCFI) of less than 0.01 or a change in RMSEA (ΔRMSEA) of less than 0.015. We used CFA with covariates to investigate the combined effect of sex, language, and grade levels on basic number skills.

To calculate the variance ratios, we used the standard scores by grade levels and then summed up the results over the grade levels. The variance ratio was calculated by dividing the male standard deviation with the female standard deviation. A larger value indicates a larger male variance.

To estimate the ratio of males and females at the ends of the distribution, low and high performers, we transformed the standard scores into Stanines (standard nine). We used the lowest and highest stanine values (1, 9) as low- and high-performance criteria. This procedure leads to groups of approximately four percentiles at both ends of the distribution.

RESULTS

Outliers and Reliability

In the tasks where the item reaction time was used to calculate the score (Number Comparison, Digit Dot Matching, Number Series), we used three steps to clean the data. First, based on eyeballing the data, extremely long response times were deleted manually as they would have had a large impact on the mean and standard deviation of the items (e.g., there were few cases where for an unknown reason the subject had stopped answering and the response to an item was over a minute). After this, values above three standard deviations of the mean were excluded. Similarly, values under 350 ms were considered unrealistic response times and were excluded from the analyses.

The second step was to clean the cases based on accuracy. In Number Comparison and Digit Dot Matching tasks, cases with the number of correct answers within the binomial probability of guessing ($p<0.05$; less than 65% correct) were removed from the analysis.

The Number Series task and the three calculation tasks had an open answer field; therefore, a different procedure to remove cases was used. Cases with less than two correct answers were removed from further analysis because we could not confirm that the subject would have tried to answer the items. The reliability of the tasks was investigated with the Spearman-Brown and Guttman split-half coefficients (split-half reliability), where a value over 0.7 indicates adequate internal consistency. The descriptives are presented in **Table 2**. More detailed information about the performances by gender and language groups is presented in **Supplementary Material**.

FUNA-DB Factor Structure

The analyses started with an investigation of the factor structure of the FUNA-DB measure. First, a one-factor model where all subtasks were set to load on a basic numerical skills factor was fitted to the data, $\chi^2$9) = 1,638.174, $p<0.001$; CFI = 0.875; TLI = 0.791; RMSEA = 0.206. This model did not fit the data very well, and modification indices indicated that the Number Comparison and Digit Dot Matching might form a separate number-processing factor while Number Series, Single-digit Addition, Single-digit Subtraction, and Multi-digit Calculations would load on a separate factor. Hence, a two-factor model with a number-processing factor and an arithmetic fluency factor was fitted to the data. This model showed good model fit and was superior compared to the one-factor model, $\Delta\chi^2(1)$ = 1,409.114, $p<0.001$; ΔCFI = 0.11; ΔRMSEA = 0.13; $\chi^2(8)$ = 229.060, $p<0.001$; CFI = 0.983; TLI = 0.968; RMSEA = 0.081.

Measurement Invariance Across Test Version, Gender, Language Group, and Grade Level

After finding the optimal factor structure, our analyses continued with multigroup CFAs to test for measurement invariance across test versions, gender, language group, and grade.

TABLE 2 | Means and standard deviations of and correlations among the basic numerical skills tasks.

Task	Girls	Boys	Total						
	M(SD)	M(SD)	M(SD)	F1.1	F1.2	F2.1	F3.1	F3.2	F3.3
F1.1 Number comparison	1,283(410)	1,246(360)	1,264(380)	1					
F1.2 Digit-dot matching	2,540(910)	2,673(900)	2,606(910)	0.721	1				
F2.1 Number series	15,977(10,640)	14,834(13,020)	15,411(11,890)	0.541	0.587	1			
F3.1 Single-digit addition	37(12.8)	39(14.7)	38(13.8)	−0.532	−0.591	−0.731	1		
F3.2 Single-digit subtraction	34(11.9)	36(13.8)	35(12.9)	−0.493	−0.555	−0.724	0.854	1	
F3.3 Multi-digit calculations	15(6.4)	16(6.9)	16(6.7)	−0.426	−0.506	−0.664	0.685	0.761	1

Note: *All correlations were significant at p<0.001.*

TABLE 3 | Summary of goodness of fit for all models used in establishing measurement invariance across test version, gender, language group, and grade level for the dyscalculia screener (FUNA-DB).

Model	χ^2	Df	p	CFI	TLI	RMSEA	ΔCFI	ΔRMSEA
Test version								
Configural invariance	284.936	40	0.0000	0.981	0.965	0.085		
Metric invariance	310.155	56	0.0000	0.980	0.974	0.074	0.001	0.011
Scalar invariance	328.474	72	0.0000	0.980	0.979	0.065	0.000	0.009
Gender								
Configural invariance	254.580	16	0.0000	0.982	0.966	0.084		
Metric invariance	268.682	20	0.0000	0.981	0.972	0.076	0.001	0.008
Scalar invariance	380.712	24	0.0000	0.973	0.966	0.084	0.008	0.008
Language group								
Configural invariance	256.378	16	0.0000	0.982	0.965	0.084		
Metric invariance	268.709	20	0.0000	0.981	0.971	0.076	0.001	0.008
Scalar invariance	287.749	24	0.0000	0.980	0.975	0.072	0.001	0.004
Grade level								
Configural invariance	247.594	56	0.0000	0.981	0.965	0.075		
Metric invariance	449.467	80	0.0000	0.964	0.953	0.087	0.017	0.012
Scalar invariance	749.418	104	0.0000	0.937	0.936	0.101	0.027	0.014

The configural model, which assumes the same factor structure but allows the factor loadings and indicator intercepts to vary across groups, was set as the baseline model in the multigroup CFAs. This model was then compared to a metric invariance (equal factor loadings) and a scalar invariance (equal factor loadings and intercepts) model. Scalar invariance was supported for test version, gender, and language group, ΔCFI<0.01; ΔRMSEA<0.015 (**Table 3**). Concerning the grade level, the metric model showed a worse model fit than the configural model in terms of ΔCFI = 0.017 but not according to ΔRMSEA<0.015. The scalar model also showed a worse model fit than the metric model in terms of ΔCFI = 0.027 but not according to ΔRMSEA<0.015. Likewise, the scalar model also showed an adequate model fit (**Table 3**). Therefore these results indicated that FUNA-DB factor scores could be compared across grades. When looking at the factor means and variances, there was a clear association with the grade level. The factor means increased with the grade level for both the number-processing factor and arithmetic fluency factor, indicating that older students had both higher number processing skills and arithmetic fluency. The variance in number-processing skills decreased when the grade level increased. The opposite pattern emerged in arithmetic fluency. It indicates that individual differences were smaller in number-processing skills and larger in arithmetic fluency in older students compared to younger students.

Relating FUNA-DB Factor Scores to Gender, Language Group, and Grade Level

Next, having established measurement invariance, the FUNA-DB number-processing factor and arithmetic fluency factor were regressed on the gender, language group, and grade level, $\chi^2(20) = 476.077$, $p<0.001$; CFI = 0.970; TLI = 0.950; RMSEA = 0.073. This model explained 37.1% of the variance in the number-processing factor and 25.9% of the variance in the arithmetic fluency factor. Girls had better number-processing skills ($\beta = 0.06$) while boys had higher arithmetic fluency ($\beta = -0.09$). Likewise, the Swedish-speaking students had better number-processing skills ($\beta = 0.11$) and arithmetic fluency ($\beta = 0.08$). As expected, the grade level had the strongest relations to the number-processing factor ($\beta = 0.63$) and arithmetic fluency factor ($\beta = 0.51$), indicating that older students had higher scores in number-processing and arithmetic fluency tasks.

To probe for possible interaction effects between the gender, language group, and grade level, a model including interaction terms was fitted to the data, $\chi^2(32) = 498.873$, $p<0.001$; CFI = 0.969; TLI = 0.951; RMSEA = 0.059 (**Figure 2**). This model explained 37.7% of

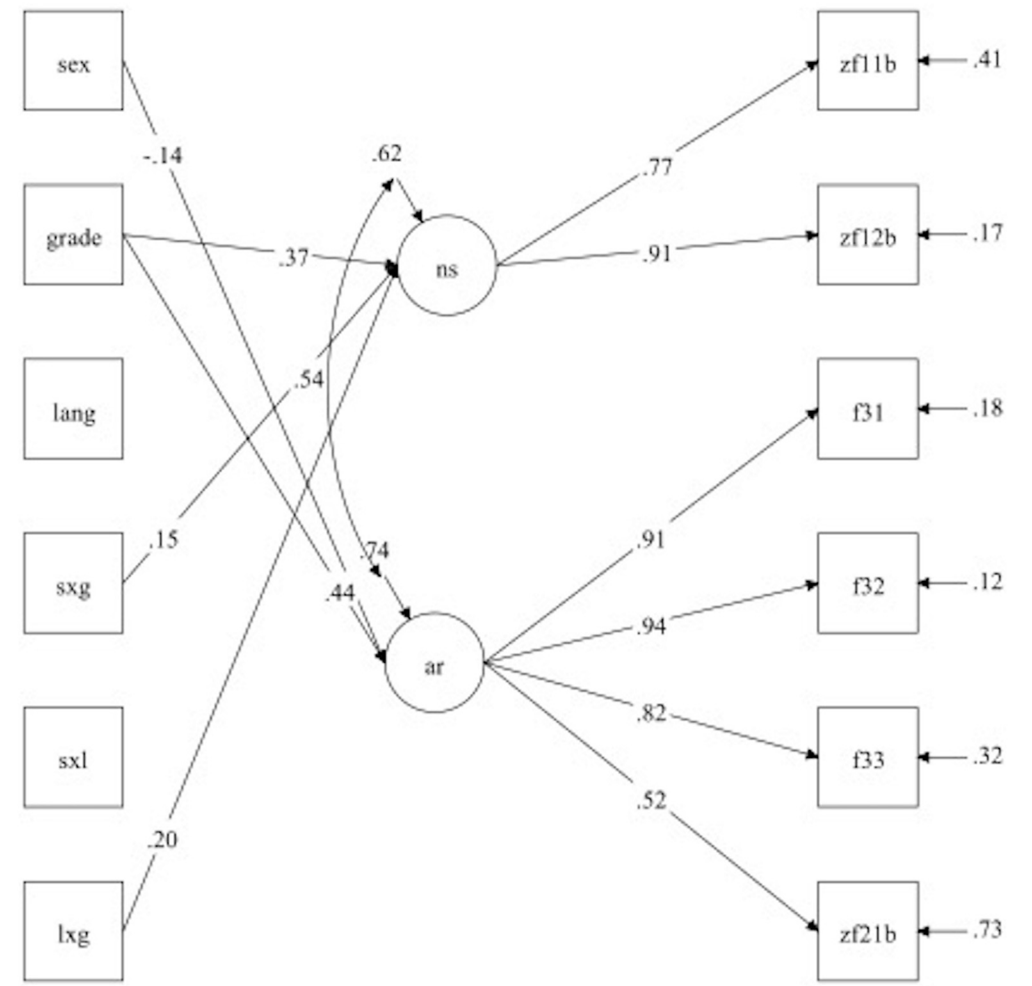


FIGURE 2 | Predicting number processing and arithmetic skills with gender, language group, and grade level. Note. ns = number processing; ar = arithmetic skills; sex = gender; grade = grade level; lang = language group; sxg = sex x grade level; sxl = sex x language group; lxg = language group x grade level; zf11b = number comparison; zf12b = dot enumeration; f31 = single-digit addition; f32 = single-digit subtraction; f33 = multi-digt addition and subtraction; zf21b = arithmetic reasoning.

the variance in the number-processing factor and 26.0% of the variance in the arithmetic fluency factor. Gender and language groups were no longer significant predictors of number processing, but the interaction gender x grade level (β = 0.15) and language group x grade level (β = 0.20) were significant. As shown in **Figure 3A**, the gender difference in favor of girls increased by the grade level. Likewise, the difference between language groups in favor of Swedish-speaking students increased by the grade level (**Figure 3B**). Concerning arithmetic fluency, the gender and grade level were the only significant predictors, not the interaction effects.

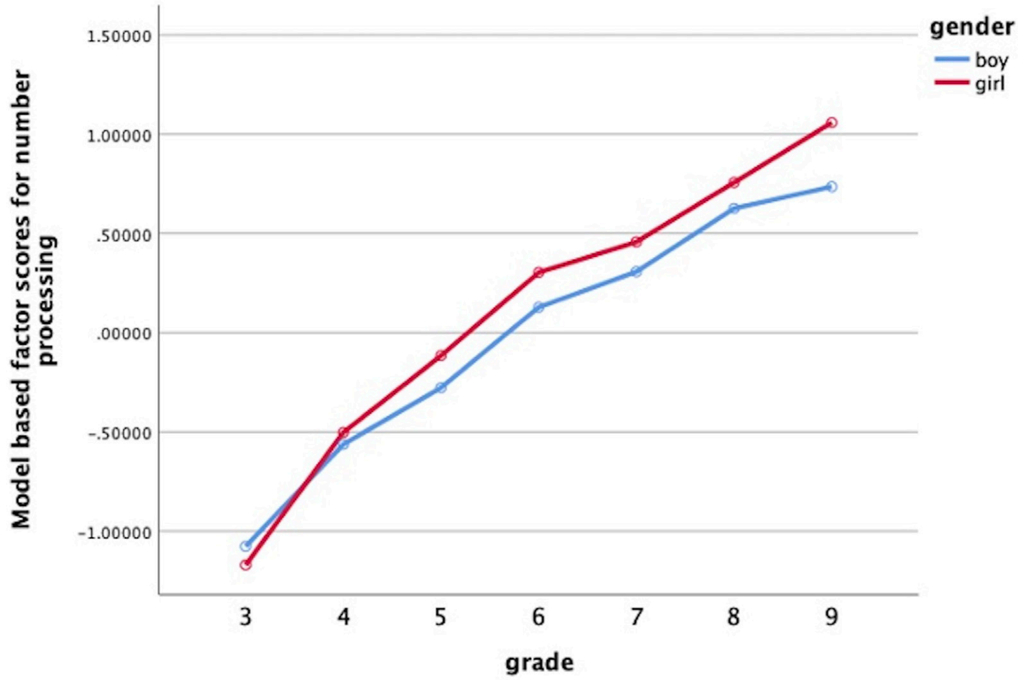
Variance Ratio and Gender Differences in the Groups With Extreme Values

We calculated standard scores for each grade level separately to analyze the means, variance, and variance ratio. The standardized means and variances are presented in **Table 4**.

There were systematic differences in arithmetic fluency tasks between the genders. First, boys performed better than girls (all $p< 0.001$), even though the effect size of this difference was small. Second, boys had a larger variance than girls, indicated by the variance ratios above VR>1.10 in all arithmetic fluency tasks (variance ratios for each task at each grade level are presented in the **Supplementary Material**).

The number-processing tasks behaved differently. In both the number comparison task and the digit-dot equivalence matching task, there was no systematic gender difference in the variance ratio. In the number comparison task, there was a small difference in average performance favoring boys (p. = 005), but the effect size of this difference was extremely small. The digit-dot equivalence matching task was the only task where girls performed better than boys ($p< 0.001$) (**Table 4**).

Last we looked at the gender ratios in the extreme groups. The groups were formed using the extreme Stanine groups 1 and 9,



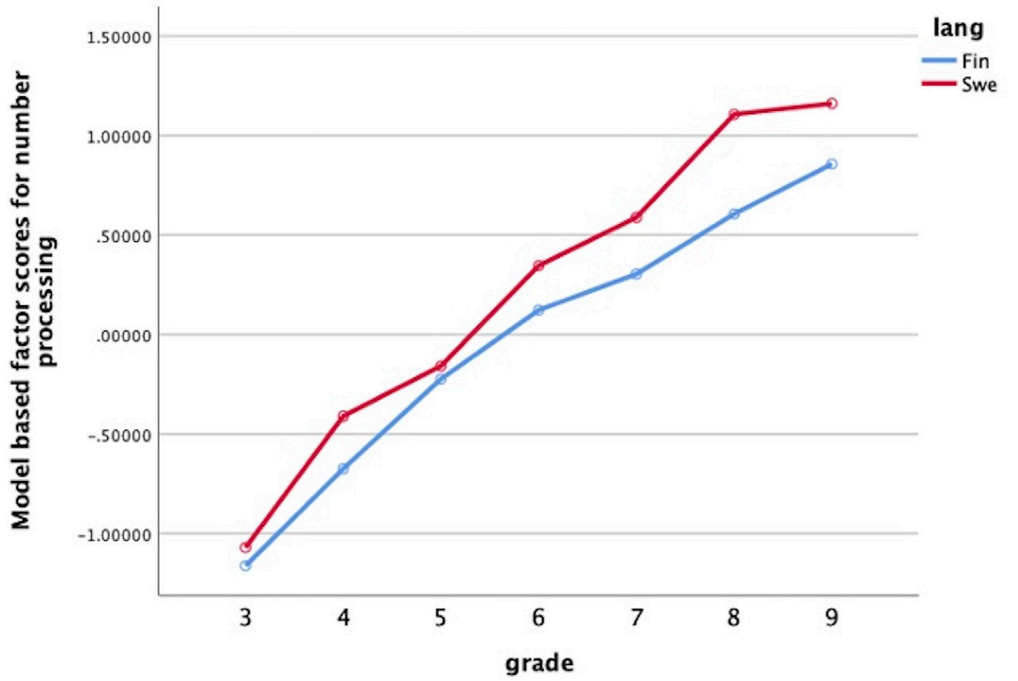


FIGURE 3 | (A) The two-way interaction between gender and grade level on number processing. lang = language group, **(B)** The two-way interaction between language group and grade level on number processing. Note. lang = language group; Fin = Finnish-speaking students; Swe = Swedish-speaking students.

each compromising about 4 percent from the end of the distribution. In all tasks measuring Arithmetic fluency, we can find more boys than girls in the groups or very low performing as well as very high-performing pupils (**Table 5**), replicating the "male variance hypothesis" (all Chi-squared <0.05). However, the number-processing tasks behaved differently. In the Number Comparison task, we find more girls than boys in the group of low performers, and in the digit-dot matching task, there are more girls in the upper end of the skill distribution (all Chi-squared <0.05). Adding language into the subgrouping did not affect the results.

TABLE4 | The standardized means, standard deviations, and variance ratios (VR) in all tasks.

Task	M		F	Eta squared	Sd		VR
	Boy	Girl			Boy	Girl	
Number comparison	0.04	−0.04	6.29*	0.002	0.97	1.03	0.94
Digit-dot matching	−0.11	0.11	35.57***	0.013	0.99	1.00	0.99
Number series	0.17	−0.17	95.62***	0.03	1.08	0.89	1.22
Single-digit addition	0.12	−0.12	41.08***	0.014	1.08	0.90	1.20
Single-digit subtraction	0.12	−0.12	36.29***	0.014	1.08	0.90	1.19
Multi-digit calculations	0.11	−0.10	23.90***	0.011	1.04	0.95	1.10

Notes. VR = Variance ratio.
**p<0.05. **p<0.01. ***p<0.001.*

TABLE5 | Percentages of subjects by gender in the low- and high-performing groups.

Tasks	Low performers		High performers	
	Boys	Girls	Boys	Girls
	%	%	%	%
Number comparison	40.1	59.9	57.3	42.7
Digit-dot matching	59.5	40.5	41.6	58.4
Number series	52.1	47.9	74.1	25.9
Single-digit addition	63.4	36.6	69.6	30.4
Single-digit subtraction	55.0	45.0	73.3	26.7
Multi-digit calculations	53.3	46.7	63.8	36.2

DISCUSSION

The present study is the first to investigate both gender and language differences at the same time in basic number skills in a large sample and with a large age range of school-aged children. Our results showed a linear development trend in basic number skills from third to ninth grade (9–15 years old in Finland). The tasks we had selected into the test battery FUNA-DB displayed good reliability and validity evidence across grade levels. A two-factor model built from number-processing skills and arithmetic fluency was found to be invariant across test versions, gender, language groups, and grade levels, and all subtasks displayed good split-half reliability.

A two-factor model suited the data better than a one-factor of numerical skills. The subtasks Number Comparison and Digit-dot Matching loaded on a number-processing factor, and the arithmetic subtasks including a numerical reasoning task (Number series) loaded on an arithmetic fluency factor. This finding is in line with existing developmental models of mathematical skills (e.g., Krajewski and Schneider, 2009; Aunio andRäsänen, 2016; Braeuning et al., 2020) that differentiates between arithmetic skills and more basic number-processing skills.

Furthermore, these basic skills are critical indicators for mathematical learning difficulties in both younger and older children (De Smedt et al., 2013; Zhang et al., 2017). The fact that our measure was found invariant across grade levels (grades 3–9) lends support to the view that students with MLD, regardless of the grade level, have problems with these basic numerical skills. Moreover, this finding and the reliability evidence indicate that the tasks selected for the assessment can be used to evaluate basic number skills across grade levels from 3 to 9.

We found that both these basic number skills showed a linear developmental trend across cohorts from grade 3 to grade 9. Concerning arithmetic fluency, this is expected as students use and train these skills during regular math classes. The age-related improvements in number-processing skills from grade 3 to grade 9 extend the finding of Brankaer et al. (2017). They observed similar changes in their numerical magnitude comparison measure from grade 1 to grade 6. It could imply two things. First, it might mean that the precision of the neurocognitive system for numerical representations matures at least till the late teenage years. Similar results have been reported in the same age range concerning the development of nonsymbolic magnitude comparison (Halberda et al., 2012). Second, it could indicate that the relationship between number processing and more advanced mathematics content might be more reciprocal than previously expected. The relationship would not be unidirectional where more advanced mathematical skills are built on basic number skills, but that practice on curriculum-based mathematics would also affect your fluency in very basic number processing leading to linear development in basic skills from early years at least to the upper primary grades.

The observed increase in variance with the grade level has also been shown in previous studies (Aunio et al., 2004; Zhang et al., 2017). An increase in variance from one grade level to another means that the difference between low- and high-performing students increases from one year to another. This kind of "Matthew effect" has been often discussed in mathematics. However, our results showed that this effect is at least partly task-dependent phenomena. We did not find a similar increase in number processing as was found in arithmetic fluency.

The second focus of our study was on looking at the gender effects on the developmental trends in basic number skills using large cross-sectional data. In our results, gender was differentially related to number-processing skills and arithmetic fluency. In number-processing skills, there was an increasing difference between genders favoring girls and Swedish-speaking pupils. Therefore, our results with tasks measuring number processing were more in accordance with the results of the mathematical achievement studies (**Figure 1**).

A systematic, but weak language-effect has been found in dual-digit comparison tasks (Nuerk et al., 2005). Moeller et al. (2015) showed that in dual-digit Arabic number comparison task, there is a systematic effect how the verbal structure of naming the numbers affects processing them. Finnish and Swedish share the same decade-unit structure in their verbal number system. Likewise, Pletzer et al. (2013) have shown a small gender difference in dual-digit comparison tasks, based on gender differences in global/local strategies. Adding a dual-digit comparison task into our battery would make it stronger to identify these effects in studies with multiple languages, and especially between languages with different structures of verbal number systems (e.g. Finnish vs. German). However, our number-processing tasks used only one-digit numbers. The Swedish number words are slightly shorter than Finnish words, but if that would have produced an effect, then there should have been a systematic difference from the early grades. We found a systematically increasing difference between the language groups in basic number processing, supporting our speculation that the language differences here reflect more cultural than cognitive effects.

However, the arithmetic fluency factor showed a different trend. Irrespective of grade and language, boys performed systematically better irrespective of the task measuring arithmetic fluency. We could not replicate Hutchison et al. (2019) results that gender similarity would be the dominating feature of the basic number skills. We conclude that both task-dependent and culture-dependent factors are affecting the gender similarities and differences.

The question of the reciprocal relationship between different basic number skills is interesting. A recent longitudinal study from first to sixth grade by Vanbinst et al. (2019) found that arithmetic skills predicted symbolic numerical magnitude processing longitudinally. Despite relatively high intercorrelation, these two types of factors showed different developmental trends in our cross-sectional study. A longitudinal approach is needed to confirm that the gender and language-dependent trends found in our study are not only a reflection of this specific moment of measurement. We must remember that in studies on curriculum-based mathematics, the results on gender differences have changed dramatically from 1 decade to another.

The previous studies with smaller sets of numerical tasks, smaller range of age groups, and smaller samples have shown mixed results concerning the gender differences or gender similarities. The recent studies of Bakker et al. (2019) and Hutchison et al. (2019) claimed that there would be no gender differences in basic number skills. Our question was if we can replicate their results in a different educational culture and with a wider age range, or if their and our results would reflect more the results typically found in the curriculum-based math achievement studies. Their study was conducted in the Netherlands, where there are no significant gender differences in mathematical skills at school age. Our study was conducted in Finland, where there has been a recent trend toward girls and especially Swedish-speaking girls performing better in mathematics than the other groups. Interestingly, only the number-processing factor seemed to follow similar trends as the more curriculum-based mathematical assessments. More direct studies are needed to assess the extent of reciprocity between the development of basic number skills and mathematical skills.

Like the developmental trends, the boy/girl variance ratios and ratios of girls vs. boys at the end of the distributions differed in the two factors. The tasks in the Arithmetic fluency factor followed the typical "male variance hypothesis," showing larger variance for boys than for girls. These values are very close to those presented by Nowell and Hedges (1998) in their analysis of gender variance from the dataset extending fifty years back. Our study is in line with the findings that even though the differences in means between the genders have mostly vanished during the last decades, the differences in variances have not. However, we found that this is also task-dependent because we did not find gender differences in variance in basic number-processing tasks (Number Comparison and Digit-dot Matching task).

Last we looked at gender differences in extremes via analyzing the gender ratios in the groups of low and high performers. We defined a pupil as a high or low performer if they belonged to the lowest or highest Stanine (standard nine) group. That means approximately four percent from both ends of the distribution. Reigosa-Crespo et al. (2012), in their large sample with similar skill factors (number processing and arithmetic), found a different ratio of boys and girls in low-performing pupils in tasks measuring number-processing skills (Number Comparison and Digit-dot Matching tasks in our study). In their study, there was twice the number of low-performing boys as girls, but they did not find any gender differences in the group of high performers. Our results did not fully replicate those results. In five out of six tasks, we found a significant overrepresentation of boys in the group of high performers (stanine class 9). Only in the Digit-dot Matching equivalence tasks, more girls showed high performance than boys. Similarly to Reigosa-Crespo's study, there were significantly more boys at the lower end of the distribution (stanine class 1) in four of the six tasks in our sample. However, in our study, the gender difference in the low-performing group was not as marked as among the high performers. As an exception, there were more girls than boys in the group of low performers in the number comparison task.

Boys were overrepresented at both ends of the distribution in most of our tasks. It was especially clear in arithmetic tasks. Depending on the arithmetic task, there were 1.09–1.73 times more boys than girls in low performers and 1.75–2.86 times more boys in the high-performing group. These numbers are close to those Nowell and Hedges (1998) reported from NAEP and other sizeable national level samples from the United States.

Limitations and Implications for Research and Practice

Although our measure displayed reliability and validity evidence, several limitations need to be considered when interpreting the results. First, our findings are based on cross-sectional data, and therefore we could not investigate the test–retest reliability of our measure. One important criterion for MLD is persistent low performance in mathematics (Mazzocco and Räsänen, 2013). With a longitudinal design, we could investigate the stability of MLD status with our measure. Second, we did not include other measures of mathematical skills to establish convergent validity. It would also have allowed us to see if the same children would be identified as at-risk for MLD with different math measures. Even when considering these shortcomings, our study adds to the literature by showing that it is possible to measure basic numerical skills with the same tasks across a broad age span. There seems to be a linear developmental trend in basic numerical skills from grade 3 to grade 9. Future longitudinal studies are needed to see if our results on increasing gender differences in number processing can be replicated in our and other educational cultures and if the relationships within and between basic number skills and curriculum-based math skills are reciprocal, as our data indicate.

We can only speculate on why we found an increase in girl advantage in number-processing skills by grade levels as our cross-sectional data did not allow for predictions over the grade level. One possible explanation could be that because 15-year-old girls in Finland outperform boys in curriculum-based mathematics (TIMMS 2018), this advantage would positively affect number processing (see Vanbinst et al., 2019 for a similar mechanism concerning arithmetic and basic number processing). This explanation would also fit the increasing advantage for Swedish-speaking pupils in number processing compared to Finnish-speaking students (TIMMS, 2018). However, our results with tasks measuring arithmetic fluency did not support the reciprocal development hypothesis. The results of the arithmetic fluency tasks were more in accordance with the theories of "male advantage" and "male variance hypothesis." Additional studies are needed to analyze if domain-general cognitive factors (spatial skills, verbal fluency) could partly explain the differences in results from one task to another.

There are several practical implications from our study. The validity information and the linear developmental trend indicate that it would be possible to use the same measure as a screener across several grade levels. This is important if we want to measure the development of the pupils objectively from one grade level to another. This kind of measure makes it easier for educators to conduct systematic screening for students at-risk for MLD and follow their development. The measure might also be suitable to assess the effects of interventions for students with special needs in mathematics education. Future studies will show how well these tasks suit repeated measurements in the context of intervention effectiveness studies.

Several findings in our study were task-dependent: the trends of development, gender differences, and the gender ratios among the low and high performers. Even though this kind of findings makes it difficult to build one theoretically meaningful interpretation of the results, it informs the researchers of numerical cognition about a crucial detail: individual and group differences may be hidden if we use summary scores of multiple variables. Developmental and cognitive factors and effects from educational practices and cultural factors may differently affect different numerical tasks. More studies analyzing the development of skills in basic number processing with different types of tasks are needed.

Finally, our study showed that even within a very homogenous and equality-nurturing culture such as Finland, we can find effects from gender and language. The language effect is fascinating because the tasks used only Arabic numbers, mathematical symbols, and dot patterns as stimuli. Luckily, the online format of the test battery allows us to build collaboration for cross-cultural studies between different countries and educational cultures easily.

The validity and reliability data of the pilot study indicate that we have good grounds to continue the development of the online FUNA-DB battery to be used as a tool to detect individual differences in basic number skills in the age group from 9 to 15 years. Future studies will show how well the battery suits differentiating low performance from specific learning disabilities (Mazzocco andRäsänen, 2013) and whether our tasks are sensitive enough to detect intervention effectiveness. The pilot study results encourage us to continue to construct assessment tools that can build a bridge between empirical research and educational practice.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Municipal research committees. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

PR and JK contributed equally to the task development, designing the tasks and the data collecting, data analysis and writing of the manuscript, PA, AL, AH, and EV participated in project work and writing the manuscript, JF participated in the data analysis, TR and M-JL participated in the task design, building the online assessments and building the datasets.

FUNDING

This research was supported by a grant from the Swedish Cultural Foundation in Finland (140884) to the research project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2021.683672/full#supplementary-material

REFERENCES

Anastasi, A. (1958). *Differential Psychology*. 3rd ed. New York, NY: Macmillan.

Aunio, P., Ee, J., Lim, S. E. A., Hautamäki, J., Luit, V., and Johannes, E. H. (2004). Young Children's Number Sense in Finland, Hong Kong and Singapore. *Int. J. Early Years Educ.* 12, 196–216. doi:10.1080/0966976042000268681

Aunio, P., Niemivirta, M., Hautamäki, J., Van Luit, J. E., Shi, J., and Zhang, M. (2006). Young Children's Number Sense in China and Finland. *Scand. J. Educ. Res.* 50 (5), 483–502. doi:10.1080/00313830600953576

Aunio, P., and Räsänen, P. (2016). Core Numerical Skills for Learning Mathematics in Children Aged Five to Eight Years – a Working Model for Educators. *Eur. Early Child. Educ. Res. J.* 24 (5), 684–704. doi:10.1080/1350293X.2014.996424

Badian, N. A. (1983). Dyscalculia and Nonverbal Disorders of Learning. *Prog. Learn. disabilities* 5, 235–264.

Bakker, M., Torbeyns, J., Wijns, N., Verschaffel, L., and De Smedt, B. (2019). Gender equality in 4-to 5-year-old Preschoolers'Early Numerical Competencies. *Develop. Sci.* 22 (1), e12718. doi:10.1111/desc.12718

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, 1976–82, Rochester, Minn. *Ambul. Pediatr.* 5 (5), 281–289. doi:10.1367/A04-209R.1

Bartelet, D., Ansari, D., Vaessen, A., and Blomert, L. (2014). Cognitive Subtypes of Mathematics Learning Difficulties in Primary Education. *Res. Develop. Disabilities* 35, 657–670. doi:10.1016/j.ridd.2013.12.010

Blume, F., Dresler, T., Gawrilow, C., Ehlis, A. C., Goellner, R., and Moeller, K. (2021). Examining the Relevance of Basic Numerical Skills for Mathematical Achievement in Secondary School Using a Within-Task Assessment Approach. *Acta Psychologica* 215, 103289. doi:10.1016/j.actpsy.2021.103289

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Braeuning, D., Ribner, A., Moeller, K., and Blair, C. (2020). The Multifactorial Nature of Early Numeracy and its Stability. *Front. Psychol.* 11, 518981. doi:10.3389/fpsyg.2020.518981

Brankaer, C., Ghesquiére, P., and De Smedt, B. (2017). Symbolic Magnitude Processing in Elementary School Children: A Group-Administered Paper-And-Pencil Measure (SYMPTest). *Behav. Res. Methods* 49, 1361–1373. doi:10.3758/s13428-016-0792-3

Bull, R., Cleland, A. A., and Mitchell, T. (2013). Gender Differences in the Spatial Representation of Number. *J. Exp. Psychol. Gen.* 142, 181–192. doi:10.1037/a0028387

Butterworth, B. (2005). The Development of Arithmetical Abilities. *J.Child Psychol.Psychiatry* 46 (1), 3–18. doi:10.1111/j.1469-7610.2004.00374.x

Committee for Economic Development of Australia [CEDA] (2013). Women in Leadership: Understanding the Gender gap. CEDA, Available at: .http://adminpanel.ceda.com.au/folders/Service/Files/Documents/15355~cedawiljune%202013final.pdf

Conole, G., and Warburton, B. (2005). A Review of Computer-Assisted Assessment. *ALT-J* 13 (1), 17–31. doi:10.1080/0968776042000339772

De Smedt, B., Noël, M. P., Gilmore, C., and Ansari, D. (2013). How Do Symbolic and Non-symbolic Numerical Magnitude Processing Skills Relate to Individual Differences in Children's Mathematical Skills? AReview of Evidence from Brain and Behavior. *Trends Neurosci. Educ.* 2 (2), 48–55. doi:10.1016/j.tine.2013.06.001

Devine, A., Soltész, F., Nobes, A., Goswami, U., and Szűcs, D. (2013). Gender Differences in Developmental Dyscalculia Depend on Diagnostic Criteria. *Learn. Instruction* 27, 31–39. doi:10.1016/j.learninstruc.2013.02.004

Dirks, E., Spyer, G., van Lieshout, E. C., and de Sonneville, L. (2008).Prevalence of Combined reading and Arithmetic Disabilities. *J.Learn. Disabilities* 41 (5), 460–473. doi:10.1177/0022219408321128

Feingold, A. (1992). Sex Differences in Variability in Intellectual Abilities: A New Look at an Old Controversy. *Rev. Educ. Res.* 62 (1), 61–84. doi:10.3102/00346543062001061

Fennema, E. (1974). Mathematics Learning and the Genders. *J. Res. Maths. Educ.* 5, 126–129. doi:10.2307/748949

Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996). Development of Arithmetical Competencies in Chinese and American Children: Influence of Age, Language, and Schooling. *Child. Develop.* 67 (5), 2022–2044. doi:10.1111/j.1467-8624.1996.tb01841.x

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *Am. Econ. Rev.* 104 (4), 1091–1119. doi:10.1257/aer.104.4.1091

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. *Develop. Psychol.* 48, 1229. doi:10.1037/a0027433

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number Sense across the Lifespan as Revealed by aMassive Internet-Based Sample. *Proc. Natl. Acad. Sci.* 109 (28), 11116–11120. doi:10.1073/pnas.1200196109

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., and Gernsbacher, M. A. (2007). The Science of Gender Differences in Science and Mathematics. *Psychol. Sci. Public Interest* 8, 1–51. doi:10.1111/j.1529-1006.2007.00032.x

Halpern, D. F. (1986). *Gender Differences in Cognitive Abilities*. Hillsdale, NJ: Erlbaum.

Hellstrand, H., Korhonen, J., Räsänen, P., Linnanmäki, K., and Aunio, P. (2020). Reliability and Validity Evidence of the Early Numeracy Test for Identifying Children at Risk for Mathematical Learning Difficulties. *Int. J. Educ. Res.* 102, 101580. doi:10.1016/j.ijer.2020.101580

Hutchison, J. E., Lyons, I. M., and Ansari, D. (2019). More Similar ThanDifferent: Gender Differences in Children's Basic Numerical Skills Are the Exception Not the Rule. *Child. Develop.* 90 (1), e66–e79. doi:10.1111/cdev.13044

Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychol. Bull.* 107 (2), 139–155. doi:10.1037/0033-2909.107.2.139

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender Similarities Characterize Math Performance. *Science* 321, 494–495. doi:10.1126/science.1160364

Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Develop. Psychol.* 45 (3), 850–867. doi:10.1037/a0014939

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., and Cantlon, J. F. (2018). No Intrinsic Gender Differences in Children's Earliest Numerical Abilities. *npj Sci. Learn.* 3 (1), 1–10. doi:10.1038/s41539-018-0028-7

Koumoula, A., Tsironi, V., Stamouli, V., Bardani, I., Siapati, S., Graham, A., et al. (2004). An Epidemiological Study of Number Processing and Mental Calculation in Greek School children. *J. Learn. Disabilities* 37 (5), 377–388. doi:10.1177/00222194040370050201

Krajewski, K., and Schneider, W. (2009). Early Development of Quantity to Number-word Linkage as a Precursor of Mathematical School Achievement and Mathematical Difficulties: Findings from a Four-Year Longitudinal Study. *Learn. Instruction* 19, 513–526. doi:10.1016/j.learninstruc.2008.10.002

Krinzinger, H., Wood, G., and Willmes, K. (2012). What Accounts for Individual and Gender Differences in the Multi-Digit Number Processing of Primary School Children? *Z. für Psychol.* 220 (2), 78–89. doi:10.1027/2151-2604/a000099

Kupari, P., Vettenranta, J., and Nissinen, K. (2012). *Oppijalähtöistä PedagogiikkaaEtsimään. KahdeksannenLuokanOppilaidenMatematiikan JaLuonnontieteidenOsaaminen*. Koulutuksentutkimuslaitos: Jyväskylänyliopistopaino. Available at: https://ktl.jyu.fi/vanhat/julkaisut/julkaisuluettelo/julkaisut/2012/d106.

Kupiainen, S., Hautamäki, J., and Karjalainen, T. (2009). *The Finnish Education System and PISA*. Helsinki, Finland: Ministry of Education. Available at: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75640/opm46.pdf.

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., and Ratliff, K. (2016). Gender Differences in Spatial Cognition: Advancing the Conversation. *Wiley Interdiscip. Rev.* Cogn. Sci. 7, 127–155. doi:10.1002/wcs.1380

Lewis, C., Hitch, G. J., and Walker, P. (1994). The Prevalence of Specific Arithmetic Difficulties and Specific reading Difficulties in 9-to 10-year-old Boys and Girls. *J.Child Psychol. Psychiatry* 35 (2), 283–292. doi:10.1111/j.1469-7610.1994.tb01162.x

Li, Y., Zhang, M., Chen, Y., Deng, Z., Zhu, X., and Yan, S. (2018). Children's Non-symbolic and Symbolic Numerical Representations and Their Associations with Mathematical Ability. *Front. Psychol.* 9, 1035. doi:10.3389/fpsyg.2018.01035

Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New Trends in Gender and Mathematics Performance: AMeta-Analysis. *Psychol. Bull.* 136, 1123–1135. doi:10.1037/a0021276

Lyons, I. M., Nuerk, H., and Ansari, D. (2015). Rethinking the Implications of Numerical Ratio Effects for Understanding the Development of Representational Precision and Numerical Processing across Formats. *J. Exp. Psychol. Gen.* 144, 1021–1035. doi:10.1037/xge0000094

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical Predictors of Arithmetic success in Grades 1–6. *Develop. Sci.* 17, 714–726. doi:10.1111/desc.12152

Maccoby, E. E., and Jacklin, C. N. (1974). *The Psychology of Sex Differences*. Stanford: Stanford University Press.

Mazzocco, M. M., and Myers, G. F. (2003). Complexities in Identifying and Defining Mathematics Learning Disability in the Primary School-AgeYears. *Ann.dyslexia* 53 (1), 218–253.

Mazzocco, M. M., and Räsänen, P. (2013). Contributions of Longitudinal Studies to Evolving Definitions and Knowledge of Developmental Dyscalculia.*Trends Neuroscience Education* 2 (2), 65–73. doi:10.1016/j.tine.2013.05.001

Moeller, K., Shaki, S., Göbel, S. M., and Nuerk, H. C. (2015). Language Influences Number Processing–AQuadrilingualStudy. *Cognition* 136, 150–155. doi:10.1016/j.cognition.2014.11.003

Molnár, G., and Csapó, B. (2019). "Technology-based Diagnostic Assessments for Identifying Early Mathematical Learning Difficulties," in *In International Handbook of Mathematical Learning Difficulties*. Editors A-M. Fritz, V. Haase, and P. Räsänen (Cham: Springer), 683–707.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., and Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center. Available at: https://timssandpirls.bc.edu/timss2019/international-results/

Nowell, A., and Hedges, L. V. (1998). Trends in Gender Differences in Academic Achievement from 1960 to 1994: An Analysis of Differences in Mean, Variance, and Extreme Scores. *Sex roles* 39 (1), 21–43. doi:10.1023/A:1018873615316

Nuerk, H. C., Weger, U., and Willmes, K. (2005). Language Effects in Magnitude Comparison: Small, but NotIrrelevant. *Brain Lang.* 92 (3), 262–277. doi:10.1016/j.bandl.2004.06.107

O'Dea, R. E., Lagisz, M., Jennions, M. D., and Nakagawa, S. (2018). Gender Differences in Individual Variation in Academic Grades Fail to Fit Expected Patterns for STEM. *Nat. Commun.* 9 (1), 1–8. doi:10.1038/s41467-018-06292-0

OECD (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: PISA, OECD Publishing. doi:10.1787/5f07c754-en

OECD (2015). *PISA the ABC of Gender equality in Education Aptitude, Behaviour, Confidence: Aptitude, Behaviour, Confidence*. Paris: OECD Publishing. doi:10.1787/9789264229945-en

Parker, P. D., Van Zanden, B., and Parker, R. B. (2018). Girls GetSmart, Boys Get Smug: Historical Changes in Gender Differences in Math, Literacy, and Academic Social Comparison and Achievement. *Learn. Instruction* 54, 125–137. doi:10.1016/j.learninstruc.2017.09.002

Penner, A. M., and Paret, M. (2008). Gender Differences in Mathematics Achievement: Exploring the Early Grades and the Extremes. *Soc. Sci. Res.* 37 (1), 239–253. doi:10.1016/j.ssresearch.2007.06.012

Pletzer, B., Kronbichler, M., Nuerk, H. C., and Kerschbaum, H. (2013).Sex Differences in the Processing of Global vs. Local Stimulus Aspects in a Two-Digit Number Comparison Task–An fMRI Study. *PloS one* 8 (1), e53824. doi:10.1371/journal.pone.0053824

Ramaa, S., and Gowramma, I. P. (2002). A Systematic Procedure for Identifying and Classifying Children with Dyscalculia Among Primary School Children in India. *Dyslexia* 8 (2), 67–85. doi:10.1002/dys.214

Räsänen, P. (2015). "Computer-assisted interventions on basic number skills," in *The Oxford Handbook of Mathematical Cognition*. Editors R. Cohen Kadosh and A. Dowker (Oxford University Press), 745–766.

Räsänen, P., Laurillard, D., Käser, T., and von Aster, M. (2019). "Perspectives to Technology- Enhanced Learning and Teaching in Mathematical Learning Difficulties," in *International Handbook of Mathematical Learning Difficulties: From lab to classroom*. Editors A. Fritz, V. Haase, and P. Räsänen (Cham: Springer International Publishing), 733–754.

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., et al. (2012). Basic Numerical Capacities and Prevalence of Developmental Dyscalculia: the Havana Survey. *Develop. Psychol.* 48 (1), 123. doi:10.1037/a0025356

Reilly, D., Neumann, D. L., and Andrews, G. (2017). Investigating Gender Differences in Mathematics and Science: Results from the 2011 Trends in Mathematics and Science Survey. *Res. Sci. Educ.* 49 (1), 25–50. doi:10.1007/s11165-017-9630-6

Reinert, R. M., Huber, S., Nuerk, H. C., and Moeller, K. (2016). Gender Differences in Number-Line Estimation: The Role of Numerical Estimation. *Br. J. Psychol.* 108 (2), 334–350. doi:10.1111/bjop.12203

Resnick, I., Newcombe, N. S., and Jordan, N. C. (2019). "The Relation between Spatial Reasoning and Mathematical Achievement in Children with Mathematical Learning Difficulties," in *International Handbook of Mathematical Learning Difficulties*.Editors A-M. Fritz, V. Haase, and P. Räsänen (Cham: Springer), 423–435.

Rosselli, M., Ardila, A., Matute, E., and Inozemtseva, O. (2009).Gender Differences and Cognitive Correlates of Mathematical Skills in School-AgedChildren. *Child. Neuropsychol.* 15, 216–231. doi:10.1080/09297040802195205

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., and Marchant, H., III (1999). Math-fact Retrieval as the Cognitive Mechanism Underlying Gender Differences in Math TestPerformance. *Contemp. Educ. Psychol.* 24 (3), 181–266. doi:10.1006/ceps.1999.1004

Shalev, R. S., Auerbach, J., Manor, O. H. A. D., and Gross-Tsur, V. (2000). Developmental Dyscalculia: Prevalence and Prognosis. *Eur. Child Adolesc. Psychiatry* 9 (2), S58–S64. doi:10.1007/s007870070009

Shen, C., Vasilyeva, M., and Laski, E. V. (2016). Here, but NotThere: Cross-National Variability of Gender Effects in Arithmetic. *J. Exp. Child Psychol.* 146, 50–65. doi:10.1016/j.jecp.2016.01.016

Thompson, C. A., and Opfer, J. E. (2008). Costs and Benefits of Representational Change: Effects of Context on Age and Gender Differences in Symbolic Magnitude Estimation. *J. Exp. Child Psychol.*, 101, 20–51.doi:10.1016/j.jecp.2008.02.003

van Tetering, M., van der Donk, M., De Groot, R. H. M., and Jolles, J. (2019). Sex Differences in the Performance of 7–12 Year Olds on aMental Rotation Task and the Relation with Arithmetic Performance. *Front. Psychol.* 10, 107.doi:10.3389/fpsyg.2019.00107

Vanbinst, K., Ghesquière, P., and De Smedt, B. (2019). Is the Long-Term Association between Symbolic Numerical Magnitude Processing and Arithmetic Bi-directional? *J. Numer. Cogn.* 5 (3), 358–370. doi:10.5964/jnc.v5i3.202

Vettenranta, J., Hiltunen, J., Kotila, J., Lehtola, P., Nissinen, K., Puhakka, E., et al. (2020a). *PerustaidoistaVauhtiaKoulutielle: NeljännenLuokanOppilaiden MatematiikanJaLuonnontieteidenOsaaminen: KansainvälinenTIMSS 2019-tutkimusSuomessa*. Jyväskylä: Koulutuksentutkimuslaitos. Available at: https://jyx.jyu.fi/bitstream/handle/123456789/73016/1/4-TIMSS-2019%20JULKAISU.pdf.

Vettenranta, J., Hiltunen, J., Kotila, J., Lehtola, P., Nissinen, K., Puhakka, E., et al. (2020b). *TulevaisuudenAvaintaidotPuntarissa: KahdeksannenLuokanOppilaiden MatematiikanJaLuonnontieteidenOsaaminen: KansainvälinenTIMSS 2019-tutkimusSuomessa*. Jyväskylä: Koulutuksentutkimuslaitos. Available at: https://jyx.jyu.fi/bitstream/handle/123456789/73019/2/8-TIMSS-2019%20JULKAISU_eidt.pdf.

Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., and Zhou, X. (2012). Gender Differences in Children's Arithmetic Performance Are Accounted for by Gender Differences in Language Abilities. *Psychol. Sci.* 23, 320–330. doi:10.1177/0956797611427168

Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M. K., and Nurmi, J. E. (2017). Knowing, Applying, and Reasoning about Arithmetic: Roles of Domain-General and Numerical Skills in Multiple Domains of Arithmetic Learning. *Develop. Psychol.* 53 (12), 2304–2318. doi:10.1037/dev0000432

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.

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/feduc.2021.655747

Quantitative and Qualitative Differences in the Canonical and the Reverse Distance Effect and Their Selective Association With Arithmetic and Mathematical Competencies

Stephan E. Vogel[1]*, *Thomas J. Faulkenberry*[2] *and Roland H. Grabner*[1]

[1]*Educational Neuroscience, Institute of Psychology, University of Graz, Graz, Austria,* [2]*Department of Psychological Sciences, Tarleton State University, Stephenville, TX, United States*

OPEN ACCESS

Edited by:
Júlia Beatriz Lopes-Silva, Federal University of Minas Gerais, Brazil

Reviewed by:
Frank Reinhold, University of Education Freiburg, Germany
Dorian Gorgan, Technical University of Cluj-Napoca, Romania

***Correspondence:**
Stephan E. Vogel
stephan.vogel@uni-graz.at

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

Received: *19 January 2021*
Accepted: *07 July 2021*
Published: *27 July 2021*

Citation:
Vogel SE, Faulkenberry TJ and Grabner RH (2021) Quantitative and Qualitative Differences in the Canonical and the Reverse Distance Effect and Their Selective Association With Arithmetic and Mathematical Competencies. Front. Educ. 6:655747. doi: 10.3389/feduc.2021.655747

Understanding the relationship between symbolic numerical abilities and individual differences in mathematical competencies has become a central research endeavor in the last years. Evidence on this foundational relationship is often based on two behavioral signatures of numerical magnitude and numerical order processing: the *canonical* and the *reverse distance effect.* The *former* indicates faster reaction times for the comparison of numerals that are far in distance (e.g., 2 8) compared to numerals that are close in distance (e.g., 2 3). The latter indicates faster reaction times for the ordinal judgment of numerals (i.e., are numerals in ascending/descending order) that are close in distance (e.g., 2 3 4) compared to numerals that are far in distance (e.g., 2 4 6). While a substantial body of literature has reported consistent associations between the *canonical distance effect* and arithmetic abilities, rather inconsistent findings have been found for the *reverse distance effect*. Here, we tested the hypothesis that estimates of the *reverse distance effect* show qualitative differences (i.e., not all participants show a *reverse distance effect* in the expected direction) rather than quantitative differences (i.e., all individuals show a *reverse distance effect*, but to a different degree), and that inconsistent findings might be a consequence of this variation. We analyzed data from 397 adults who performed a computerized numerical comparison task, a computerized numerical order verification task (i.e., are three numerals presented in order or not), a paper pencil test of arithmetic fluency, as well as a standardized test to assess more complex forms of mathematical competencies. We found discriminatory evidence for the two distance effects. While estimates of the *canonical distance effect* showed quantitative differences, estimates of the *reverse distance effect* showed qualitative differences. Comparisons between individuals who demonstrated an effect and individuals who demonstrated no *reverse distance effect* confirmed a significant moderation on the correlation with mathematical abilities. Significantly larger effects were found in the group who showed an effect. These findings confirm that estimates of the *reverse distance effect* are subject to qualitative differences and that we need to better characterize the underlying mechanisms/ strategies that might lead to these qualitative differences.

Keywords: individual differences, canonical distance effect, reverse distance effect, arithmetic abilities, mathematical competencies

INTRODUCTION

In the past years, there has been an increase in interest to better understand the cognitive foundation of symbolic numerical abilities and its relationship to arithmetic and mathematical competencies. This upsurge has emerged from the observation that arithmetic abilities are equally important for life success as literacy (Parsons and Bynner, 2005) and that deficits in this domain can have detrimental effects on individuals wellbeing as well as on nation's economy (Gross et al., 2009). Results of this research have provided evidence that measures of two symbolic concepts are associated with arithmetic abilities: numerical magnitude (i.e., knowledge about which numeral is larger or smaller) and numerical order (i.e., knowledge about the relative rank or position of a numeral with a sequence).

The existing evidence on the relationship of these basic numerical abilities with arithmetic abilities is largely based on two behavioral signatures: The *canonical distance effect* and the *reverse distance effect*. The *canonical distance effect* emerges when participants decide as fast as possible, without making mistakes, which of two numerals is larger/smaller (e.g., 2 6). Reaction time measures of this comparison task have been shown to be inversely related to the numerical distance of the numerals (Moyer and Landauer, 1967). In other words, participants are faster when the distance between the numerals is larger (e.g., 8 2) compared to when it is smaller (e.g., 2 3). The *canonical distance effect* is a well replicated finding (e.g., De Smedt et al., 2009; Holloway and Ansari, 2009; Lonnemann et al., 2011; Sasanguie et al., 2012; Vogel et al., 2015; Goffin and Ansari, 2016) and, although still debated, it is assumed to reflect the internal representation of numerical quantities (Moyer and Landauer, 1967; for alternative explanations see; Van Opstal et al., 2008; Zorzi and Butterworth, 1999).

Individual differences of the *canonical distance effect* show a consistent negative correlation with arithmetic performance (i.e., the smaller the *canonical distance effect*, the better arithmetic performance) in children (e.g., De Smedt et al., 2009; Holloway and Ansari, 2009; Lonnemann et al., 2011; Sasanguie et al., 2012; Vogel et al., 2015) as well as in adults (e.g., Goffin and Ansari, 2016; Maloney et al., 2010). In other words, individuals who perform better in arithmetic demonstrate a smaller *canonical distance effect* compared to individuals who perform worse, possibly due to a more precise representation of symbolic numerical quantities (Holloway and Ansari, 2009). Significant differences in the size of the *canonical distance effect* have also been reported for individuals with learning difficulties (i.e., developmental dyscalculia; e.g., Ashkenazi et al., 2008; Delazer et al., 2006; Price et al., 2007; Rousselle and Noël, 2007). Together, these findings indicate a significant correlative association with arithmetic and mathematical abilities, which a meta-analysis quantified with a small effect size of $r = 0.135$[1] (Schneider et al., 2017; for a review see; De Smedt et al., 2013).

The *reverse distance effect* relates to the numerical order verification task (Franklin et al., 2009; Lyons and Beilock, 2011). In this task participants verify as fast as possible, without making mistakes, whether the order of three numerals is correct (e.g., 2 3 4) or incorrect (e.g., 3 4 2). Several studies have shown that ordinal judgment tends to be faster for adjacent numbers (e.g., 2 3 4) compared to distant numbers (e.g., 2 4 6) in the correct order condition (i.e., numbers that are in correct ascending or descending order). Because of its opposite direction, i.e., faster reaction times for small distances, the effect has been labeled as the *reversal* of the *canonical distance effect* (Turconi et al., 2006; Franklin et al., 2009; Lyons and Beilock, 2011; Lyons and Beilock, 2013). Several studies have confirmed the existence of a *reverse distance effect* in children (Lyons and Ansari, 2015; Vogel et al., 2015) as well as in adults (Franklin et al., 2009; Lyons and Beilock, 2011; Lyons and Beilock, 2013; Vogel et al., 2017; Vogel et al., 2019; Vos et al., 2017; Sella et al., 2020). And although the nature of the *reverse distance effect* is not well understood, some research indicates that it is associated with an effective retrieval mechanism of learned ordinal sequences from long-term memory (Lyons et al., 2016; Sasanguie and Vos, 2018; Vogel et al., 2019; Sella et al., 2020; Sommerauer et al., 2020). While items with larger distances might be solved via a sequential and procedural comparison process (e.g., 2 4 6 = 2 > 4 and 4 > 6), small distances (especially consecutive items) might be retrieved as sequence-lists (e.g., chunks, Dehaene et al., 2015) from long-term memory.

In contrast to the *canonical distance effect*, inconsistent findings have been reported in the few studies that have investigated the correlative association between the *reverse distance effect* and arithmetic abilities. Some studies have found a negative (i.e., the smaller the reverse distance effect, the better arithmetic performance; Goffin and Ansari, 2016), while other studies have found a positive (i.e., the larger the *reverse distance effect*, the better arithmetic performance; Vogel et al., 2019) or no relationship at all (Orrantia et al., 2019; Vogel et al., 2015; Vogel et al., 2017; Vos et al., 2017). These findings are in contrast to the consistently positive correlations reported for the *canonical distance effect*.

One possible explanation for the inconsistent findings is that the *reverse distance effect* is not a quantitative (i.e., all individuals show a reverse distance effect, but to a different degree), but rather a qualitative measure of individual differences (i.e., not all participants show a reverse distance effect in the expected direction; see also Faulkenberry and Bowman, 2020; Haaf and Rouder, 2019). More specifically, the involvement of two (or even more) strategies in the numerical order verification task could introduce combinatorial variations that lead to qualitative differences in how the task is performed. As discussed above, the processing of ordinal information has been associated with at least two different strategies: an effective retrieval of learned and automatized sequences (mainly used with small distances; e.g., 1 2 3) and a less effective sequential magnitude comparison process (mainly used with larger distances; e.g., 2 4 6). Individual variations of these strategies could result in qualitative differences (e.g., some might use magnitude comparison mechanisms more often than others) that might have obscured the correlation of the *reverse distance effect* with arithmetic and mathematical abilities in previous studies.

[1]Please note that the calculated effect size did not differentiate between the *canonical distance effect* derived from symbolic (i.e., using Arabic numerals) and non-symbolic (i.e., dot arrays) comparison tasks. Since larger correlations with mathematical abilities are typically observed with symbolic measurements (De Smedt et al., 2013; Schneider et al., 2017), a larger effect size might be expected for the symbolic *canonical distance effect*.

Evidence to answer this question is extremely sparse, since existing studies have assumed a quantitative structure for the *reverse distance effect* (e.g., Goffin and Ansari, 2016; Vogel et al., 2017; Vogel et al., 2019). The possibility of a qualitative structure has, to the best of our knowledge, not been systematically investigated or described. Nevertheless, important information can be gained from studies that have investigated both distance effects within the same individuals. For instance, (Goffin and Ansari, 2016) collected data from a sample of 68 adults. The participants performed a computerized numerical comparison task to measure the *canonical distance effect*, a numerical ordinal verification task to measure the *reverse distance effect* and a test of arithmetic performance (i.e., Woodcock Johnson III Tests of Achievement; Woodcock et al., 2001). The results of the reaction time analyses showed that the distance effect measures were uncorrelated with one another (r = 0.17, p = n.s.) and that both effects explained unique variance in their relationship with arithmetic performance (*canonical distance effect*: r = −0.310, $p < 0.05$; *reverse distance effect*: r = −0.422, $p < 0.01$). However, the authors did not assess the structure of individual differences for the *reverse* and the *canonical distance effect* in detail. Figure 2 of that study (p.73; Goffin and Ansari, 2016) indicates that some individuals did not show the expected *reverse distance effect* in the ordinal verification task but the opposite: a *canonical distance effect*. This finding contrasts the numerical comparison task in which almost all individuals showed the expected *canonical distance effect*. Thus, the result pattern indicates that estimates of the *reverse distance effect* might be subject to qualitative differences, while estimates of the *canonical distance effect* might be of quantitative nature.

In the present work we tested this hypothesis and explored whether a qualitative individual differences structure for the estimates of the *reverse distance effect* moderate the association with arithmetic and more complex forms of mathematical abilities. We used the approach of Haaf and Rouder (2017) and Haaf and Rouder (2019) to investigate the structure of individual differences for the estimates of the *reverse* and the *canonical distance effects*. Their Bayesian approach instantiates different models which place varying levels of constraint on individual differences. Key among these are two models which reflect qualitative and quantitative individual differences: an *unconstrained* model and a *positive-effects* model. The unconstrained model allows individual differences to vary among all possible values (positive or negative), and thus reflects qualitative differences. The positive-effects model assumes that all effects are positive. This model reflects quantitative differences—since all individuals show an effect in the expected direction (positive values) the only variation is in the magnitude of the effect. The common-effect model places even more constraint on individual variation by assuming that everyone's distance effect is the same value (i.e., there is an effect in the expected direction, and the size of the effect is equal across individuals). The null model is the most constrained and it specifies that the effect is zero (i.e., there is no effect across all individuals). The best model fit is then tested using a Bayes factor model comparison. This novel approach has been successfully implemented to test the structure of individual differences in numerical priming effects (Haaf and Rouder, 2019), location and color Stroop effects (Haaf and Rouder, 2019), numerical size congruity effects (Faulkenberry and Bowman, 2020), and the truth effect (Schnuerch et al., 2020).

As an alternative, clustering methods (e.g., latent profile analysis) could be used to model individual differences in the various distance effects. Such methods work by collapsing the high dimensional space of response variables into configural profiles (or clusters), allowing the analyst to classify individuals based on cluster membership. One limitation of such methods is that they do not clearly account for the qualitative distinctions between positive and negative effects. As Haaf and Rouder (2019) point out, a cluster analysis would likely place two individuals with true distance effects of −20 and 20 ms into the same cluster, whereas two individuals with true effects of 20 and 200 ms would not be placed together. We believe that the distinction between positive and negative distance effects is important, as each points to a different theoretical mechanisms of number processing. Thus, instead of using a single model (e.g., a clustering model), we compared several different models, each of which specified a different level of constraint on the true distance effects that could be present among individuals.

Using data from a group of adults who performed a computerized numerical comparison task, a numerical order verification task, a paper-pencil test of arithmetic fluency, as well as a standardized measure assessing mathematical abilities, we tested the following hypotheses: 1) Is there a *canonical* and a *reverse distance effect* on the group level? Based on a large body of evidence we expected to replicate a) significant faster reaction times for large distances compared to small distances in the numerical comparison task, and b) significant faster reaction times for small distances compared to large distances in the correct order condition of the numerical order verification task. 2) Are individual differences in the distance effects quantitative or qualitative? Based on our hypothesis described above, we expected that the best Bayesian model fit for the estimates of the *reversed distance effect* would be an unconstrained model (i.e., not all participants show a distance effect in the expected direction), while the best fit for the estimates of the *canonical distance effect* would be a positive-effect model (i.e., all individuals show a *canonical distance effect*, but to a different degree). 3) Do the model estimates of the *reverse distance effect* moderate the association with arithmetic and mathematical abilities? Based on our hypothesis, we expected that if the estimates of the *reverse distance effect* are subject to qualitative differences, the correlative association with arithmetic and mathematical abilities should be significantly larger in a selected group of individuals who truly show a *reverse distance effect,* in comparison to a group of individuals who show no evidence for a *reverse distance effect*.

METHODS

Participants

We collected behavioral data from 450 adult participants (273 females; M_{age} = 22.32; SD = 4.66, *range* = 17–50). From this data set, we removed individuals with missing data (n = 18) and

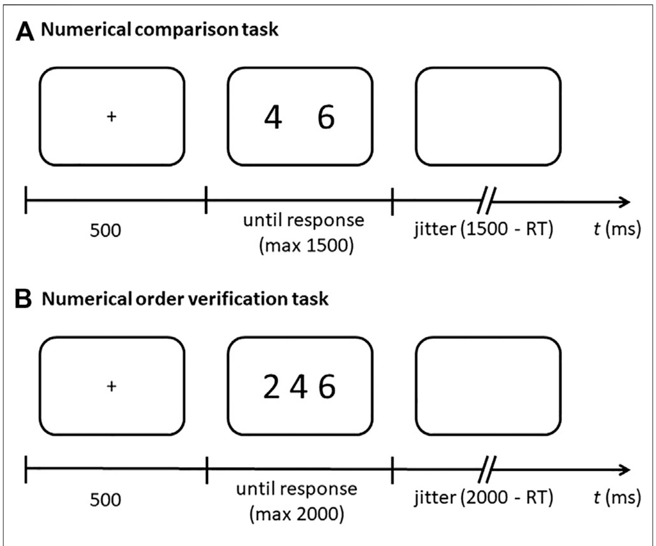


FIGURE 1 | Timing of a trial in **(A)** the numerical comparison and **(B)** the numerical order verification task. After showing a fixation cross, the numbers were presented on the screen. Participants responded *via* a button press.

individuals who reported neurological disorders and/or learning disabilities (n = 35). Thus, the final sample comprised 397 healthy participants (258 females; 354 right-handed, 29 left-handed, 14 ambidextrous) with a M_{age} of 22.32 (SD = 4.74; *range* = 17–50) years. All subsequent analyses are based on this final sample. Approximately 43% of the participants were students and reported to be enrolled in psychology, 19% in science, 18% in humanities, 11% in engineering, 5% in law or economics, and 3% without categorization. All participants gave written informed consent prior to participation and received feedback regarding their intellectual abilities after testing as incentive for taking part in the study. The local ethics board of the University of Graz approved the study.

Materials

Raw Data, analyses scripts and supplemental materials can be accessed via the open science framework (OSF) using the following link: https://osf.io/jvmc2/?view_only=64abdcbd0cac4ca0b8154ebbcc4437a2.

Numerical Comparison Task

Two single-digit numerals were horizontally presented on a computer screen (e.g., 2 8), and participants had to indicate as accurately and as fast as possible which of the two numbers is numerically larger (see **Figure 1A**; see also supplemental materials for a detailed list of the stimuli). The reason for including only single-digit numbers is that the comparison (or the ordinal verification) of two-digit numbers introduces additional reaction time effects that are not the focus of the present work (e.g., compatibility effect, decade crossing; Franklin et al., 2009; Nuerk et al., 2001). Therefore, the stimuli consisted of the Hindu-Arabic numerals 1 to 9. In half of the trials, the larger numeral was presented on the left side. In the other half, the larger numeral was presented on the right side. The numerical distance between the numerals (i.e., inter-item distance) was systematically manipulated to measure the *canonical distance effect*. We categorized trials (80 in total) into small (40 trials with a numerical distance of one and two: e.g., 2 3; 5 3) and large inter-item distance trials (40 trials with a numerical distance of five and six: e.g., 2 7; 8 2).

The presentation of the stimuli started with a fixation (500 ms), then the two numerals were simultaneously presented until a key response was given (maximum presentation time of the two numerals 1500 ms), followed by a blank screen with a variable jitter (calculated as the difference between 1500 ms and the response time of the trial). Reaction time data were recorded to estimate individual's *canonical distance effect* for correct trials.

Numerical Order Verification Task

This task was adapted from Vogel et al. (2017) and Vogel et al. (2019). Three single-digit Arabic numerals were horizontally presented on a computer screen (see **Figure 1B**; see also **Supplementary Material** for a detailed list of the stimuli) and participants had to evaluate, as accurately and as fast as possible, whether the three numbers represent a correct (e.g., 2 3 4) or incorrect numerical order (e.g., 2 4 3). Again, only single-digit numbers were used to avoid additional reaction time effects. The stimuli consisted of the Hindu-Arabic numerals 1 to 9. In half of the trials, the numerals were arranged in a correct ascending/descending order (e.g., 2 3 4; 6 5 4). In the other half, the numerals were arranged in an incorrect mixed order (e.g., 2 4 3; 4 2 3). Again, the inter-item distance was manipulated in order to measure the reverse distance effect. We categorized items into small (30 trials with a numerical distance of one: e.g., 2 3 4) and large distance trials (30 trials with a numerical distance of two or three: e.g., 2 4 6, 2 5 8).

Stimuli presentation started with a fixation cross (500 ms), then the three numerals were simultaneously presented on the screen until a key response was given (maximum presentation time of 2000 ms). A blank screen with a variable jitter (calculated as the difference between 2000 ms–response time) was presented at the end of each trial. We recorded reaction time data to estimate the *reverse distance effect* for the correct order condition.

Arithmetic Fluency

We assessed arithmetic performance with a paper-pencil task designed in our laboratory (Schillinger et al., 2018; Vogel et al., 2017; Vogel et al., 2019; the assessment with all items can be found on OSF) based on the French kit test (French et al., 1963). The task measures the ease with which individuals can solve small and large multiplications, additions, and subtractions problems.

Small problems include 64 single-digit multiplications (e.g., 5 × 7), 128 single-digit additions (e.g., 4 + 7), and 128 subtractions with a minuend between 4 and 20 and a single-digit subtrahend (e.g., 16–8). Research has shown that adults solve such simple arithmetic problems, especially multiplications and additions, by retrieving the respective solution from long-term memory (Ashcraft, 1992; Campbell and Xue, 2001; Grabner and De Smedt, 2011).

Large problems included 60 problems for each operation. Multiplications were composed of a double-digit number (smaller than 100) and a single-digit number (e.g., 39 × 5), additions required to sum up three double-digit numbers (e.g., 30 + 98 + 59), and subtractions consisted of two double-digit numbers (e.g., 82–31). Research has shown that such complex arithmetic problems usually require the application of an arithmetic procedure to be solved (Ashcraft, 1992; Campbell and Xue, 2001; Grabner and De Smedt, 2011).

In the test session, participants solved as many problems as possible on each sheet (the operations were printed on separate sheets) within a limited time (90 s for small and 120 s for large problems). Instead of a composite score, we calculated scores (number of correctly solved items) for each operation and problem sizes (i.e., small subtraction, large subtraction, small additions, large additions, small multiplications, and large multiplications).

Mathematics Test (M-PA)

We used the short version of the German mathematics test for selection of personnel (*Mathematiktest für die Personalauswahl*, M-PA; Jasper and Wagener, 2011) to asses individual differences in higher-order mathematics. The M-PA was developed to assess mathematical competencies of individuals with at least lower secondary education between the ages of 16 and 40. The short version consists of 31 mathematical problems with a multiple-choice (MC) or open answer (OA) format. Problems cover a wide range of mathematical topics including fractions (3 OA), conversion of units (3 OA), exponentiation (7 OA), division with decimals (2 OA), algebra (1 MC), geometry (1 MC), roots (7 OA), and logarithms (7 OA). Following instructions, participants had a total of 15 min to solve the problems. The short version of the M-PA has been reported to have good internal consistency (Cronbach alpha = 0.89) and to be highly correlated with the long version of the M-PA ($r = 0.93$), which contains a total of 77 items (Jasper and Wagener, 2011). We calculated the total number of correctly solved items for our statistical analysis.

Procedure

Data collection took place between 2015 and 2019 in a group testing room at the Institute of Psychology, University of Graz, as part of a larger and ongoing investigation. We tested participants in small groups (the size of each group varied from four to twelve individuals). Upon arriving, participants were seated in front of a computer screen and a test booklet. Participants worked through the test booklet and took a pause whenever they reached a page with a red stop sign. For all speeded tests (e.g., M-PA: 15 min and arithmetic fluency test: 10.5 min), our experimenters took the time and informed participants when they had to stop working on the respective test. Please note that in addition to the above-described tasks, the test booklet contained several additional assessments (e.g., tests assessing creativity and personality as well as questionnaires on math anxiety and general anxiety) that are not within the scope of the present study. At the end, participants were asked to answer demographic questions regarding sex, age, field of study, and final high school grade in mathematics.

Next, we collected the data from the computerized tasks. The computerized tasks (i.e., numerical comparison task and the numerical order verification task) were presented on a Dell computer (Windows 10 64-bit operating, Intel i5-4590 processor @ 3,3 gigahertz and 8 gigabyte ram) with the stimuli presentation software Psychopy (version 1.85.3; Peirce, 2008). Stimuli were visualized with a Samsung S24C450 monitor (24 inch) using a sampling rate of 60 Hz. Before each task, participants solved 6 practice trials in which they received a feedback on whether their response was correct or incorrect. The entire testing took about 2h and 30 min.

MODELING AND ANALYSIS

We used frequentist and Bayesian analyses to answer the questions of this project. All statistical analyses, including Bayesian modeling of individual differences, were calculated in R (R Core Team, 2020). Descriptive statistics provide cumulative information about all variables and their distributional properties, whereas inferential statistics are based on reaction time data (see also Goffin and Ansari, 2016).

Testing Distance Effects on the Group Level

First, we calculated two analysis of variance (ANOVA) for repeated measurements, including inter-item distance (distances 1, 2, 5, 6 in the numerical comparison task, and distances 1, 2, 3 in the numerical ordinal verification task) as the main factor. Greenhouse Geisser corrected estimates (Greenhouse and Geisser, 1959) are reported as the data violated the assumption of equal variance differences across the conditions (Mauchly's test of sphericity are all $p < 0.05$; Field et al., 2012). We used pairwise t-tests, corrected for multiple comparisons (false discovery rate (FDR) method; Benjamini and Hochberg, 1995), to test for significant differences between the single inter-item distance conditions. Partial-eta (η_p^2) and Cohen's d are reported as effect sizes for the ANOVA and t-tests.

Testing the Model Fit of the Distributional Properties of the Distance Effects

In a second step, we tested the hypothesis of a qualitative model of the estimates of the *reverse distance effect* (i.e., not all participants show the expected *reverse distance effect*) and the hypothesis of a quantitative model of the estimates of the *canonical distance effect* (i.e., all individuals show the expected *canonical distance effect*, but to a different degree). To investigate the structure of individual differences, we used the approach of Haaf and Rouder (2017) see also Faulkenberry and Bowman (2020) to develop and to test a set of four hierarchical Bayesian models. Thus, each of these models reflects a different underlying distributional structure of the distance effects θ_i (see **Supplemental Material** for a more detailed description of model specification):

1) The *unconstrained model* places no constraints on the individual distance effects. In this model, we allow subjects' distance effects to vary among all possible values (positive or negative), so we use this model to capture qualitative individual differences.

$$\mathcal{M}_u : \theta_i \sim Normal(\nu, \eta^2) \quad (1)$$

Here ν and η^2 represent the mean and variance, respectively, of the distribution of individual distance effects θ_i. The values of ν and η^2 are estimated from the observed data.

2) The *positive-effects model* places constraints on the distribution of the distance effects by assuming that all distance effects are positive. Thus, we use this model to capture quantitative differences.

$$\mathcal{M}_+ : \theta_i \sim Normal_+(\nu, \eta^2) \quad (2)$$

3) The *common-effect model* places even more constraint on the distribution of the distance effects by assuming that everyone's distance effect is the same value.

$$\mathcal{M}_1 : \theta_i = \nu \quad (3)$$

We note that if the common-effect model is the best predictor of our observed data, then such results would call into question the efficiency of our experimental design as a test to elicit individual differences.

4) The *null model* is the most constrained of the four, and it specifies that each participant's distance effect is zero:

$$\mathcal{M}_0 : \theta_i = 0 \quad (4)$$

In this model, any observed variation in response times would be due to sampling noise.

We then used Bayes factors (Jeffreys, 1968; Kass and Raftery, 1995) to test which of the four competing models is the best predictor of our observed data. Bayes factors index the relative predictive adequacy of two models by comparing the marginal likelihood of observed data under one model compared to another (Faulkenberry et al., 2020). For example, a Bayes factor of 5 indicates that the observed data are five times more likely under one model compared to another (see also **Supplementary Material** for a more detailed description of the procedure). To find out how much these results depended on our choice of prior specification, we also conducted a sensitivity analysis (see also **Supplementary Material** for prior specification). For this we adjusted the prior scales on the size of our expected *reverse* and *canonical distance effects*, relative to overall variability as well as on the by-subject variability of the effect, relative to overall variability.

In the case that the model comparisons reveal evidence for an unconstrained model (i.e., qualitative individual differences), we further classified individuals according to the type of distance effect individuals exhibited; that is, either positive, negative, or undecided. If at least 75% of the posterior samples for a specific θ_i were positive [i.e., $p(\theta_i > 0 \mid data) > 0.75$], we classified subject *i*'s distance effect as "positive". On the other hand, if at least 75% of the posterior samples were negative [i.e., $p(\theta_i < 0 \mid data) > 0.75$], we classified subject i's distance effect as "negative". In cases where less than 75% of the samples were positive or negative, we classified subject *i*'s distance effect as "undecided". Based on this classification we calculated the percentage of individuals who showed a "positive", a "negative" or no (undecided) distance effect. The benchmark for choosing a positive/negative classification was based on a similar classification by Schnuerch et al. (2020). Please note that a posterior probability of 0.75 equates to an odds ratio of 3-to-1, which is considered a minimum threshold of evidence in Bayesian model comparison.

Testing the Associations of the Reverse Distance Effect With Arithmetic and Mathematical Performance

Based on the above model specification, we estimated individual distance effects θ_i for each subject and distance effect. This was done by obtaining 10,000 posterior samples from the unconstrained model for the parameters θ_i. Then, the estimate $\hat{\theta}_i$ of each subject's distance effect θ_i was defined as the mean of these posterior samples.

We then used these individual estimates to explore the impact of a possible qualitative distribution on the association with arithmetic and mathematical performance measures. More specifically, we analyzed whether the size of the correlation coefficients differs as a factor of whether we include all individuals (as has been done in previous research) or a selection of individuals (individuals with evidence for a reverse distance effect) into the analysis. In the first analysis, we used the entire sample (regardless of whether individuals showed a distance effect or not) to calculate zero-order and partial correlations (controlling for age and the other distance effect) between the distance effects and our measures of arithmetic fluency (i.e., small and large subtractions, additions and multiplications) and mathematical competencies (M-PA).

We then repeated the above described correlation analyses with two selected groups: one group in which all individuals showed a "positive" reverse distance effect, and another group in which individuals showed "no-positive effect".

Finally, we tested whether the size of the observed correlation coefficients differed across these groups. In other words, we tested whether correlation coefficients (e.g., the correlation between *reverse distance effect* and small subtraction problems) in the "positive effect" group are significant larger compared to the "no-positive effect" group. We used r.test from the psych package (Revelle, 2020) to test for the significance of correlation differences between two different sample sizes. For this a z-score is calculated that finds the difference between the z transformed correlations divided by the standard error of the difference of two z-scores (Cohen et al., 2013).

$$z = \frac{z1 - z2}{\sqrt{\frac{1}{(n1-3)+(n2-3)}}} \quad (5)$$

Obtained *p*-values from all analyses were FDR corrected for multiple comparisons. This procedure enabled us to test whether the association between the reverse distance effect and arithmetic/mathematical abilities is moderated by group composition.

RESULTS

Descriptive Statistics

Tables 1, **2** depict descriptive statistics of the computerized tasks (i.e., numerical comparison and numerical order verification), the arithmetic fluency measures and the M-PA.

TABLE 1 | Descriptive statistics of mean reaction time and accuracy measures of the computerized tasks.

Numerical comparison	**distance 1**	**distance 2**	**distance 5**	**distance 6**
RT $_{\text{in ms}}$	489(69)	470(60)	429(50)	422(47)
AC $_{\text{in \% correct}}$	93.12(7.19)	95.13(6.00)	99.33(3.28)	99.48(3.16)

Ordinal verification	**In-order condition**			**Mixed-order condition**		
	distance 1	**distance 2**	**distance 3**	**distance 1**	**distance 2**	**distance 3**
RT $_{\text{in ms}}$	745(119)	760(129)	760(127)	851(137)	813(129)	780(134)
AC $_{\text{in \% correct}}$	92.88(6.43)	92.69(7.19)	93.54(6.90)	82.06(15.10)	85.68(16.82)	91.14(15.44)

Note: Standard deviations in parenthesis.

TABLE 2 | Descriptive statistics of the arithmetic measures and the M-PA.

	Min	**1st Qu**	**Median**	**Mean**	**3rd Qu**	**Max**	**Skew**	**Kurtosis**	***W***
Small problems									
Subtraction	8	37	47	49	58	106(128)	0.832	0.870	0.987[c]
Addition	10	52	63	63.68	74	116(128)	0.353	0.456	0.987[c]
Multiplication	13	29	41	41.12	52	64(64)	0.053	−1.100	0.956[b]
Large problems									
Subtraction	2	17	21	22.92	28	60(60)	0.849	1.258	0.960[c]
Addition	0	10	13	12.98	16	31(60)	0.592	1.055	0.975[c]
Multiplication	0	5	9	10.15	13	46(60)	1.497	3.216	0.960[c]
M-PA	6	17	22	21.16	26	31	−0.484	−0.491	0.963[c]

[a] *p < 0.05.*
[b] *p < 0.01.*
[c] *p < 0.001.*
Min, minimum value; 1st Qu, first quartile; 3rd Qu, third quartile; Max, maximum value (maximum possible value in parenthesis); Skew, skewness; W, *critical values of the Shapiro-Wilk test.*

Shapiro-Wilk tests indicated that all measures of the arithmetic fluency test and the scores from the M-PA differed significantly from a normal distribution. The distributions of the small problems showed the following characteristics: The subtraction scale was right-skewed with a leptokurtic distribution, the addition scale was right-skewed with a leptokurtic distribution, the multiplication scale was not skewed but showed a platykurtic distribution. The distributions of the large problems showed the following characteristics: The subtraction scale was right-skewed with a leptokurtic distribution, the addition scale was right-skewed with a leptokurtic distribution, and the multiplication scale was right-skewed with a leptokurtic distribution. The M-PA scale was left-skewed with a platykurtic distribution.

Inferential Statistics

Distance Effects on the Group Level

We calculated two ANOVAs to test the presence of distance effects on the group-level. As expected, the results of these analyses showed significant distance effects for both conditions. The ANOVA performed on the data in the numerical comparison task showed a significant effect of inter-item distance, $F(2.02, 801.39) = 967.235$, $p < 0.001$, $\eta_p^2 = 0.71$. T-test comparisons revealed significant differences across all distances (all comparisons $p_{FDR\text{-}adjusted} < 0.001$; effect sizes ranged from small, distance 5 ~ distance 6, $d = 0.420$, to large, distance 1 ~ distance 6, $d = 1.819$). This pattern is consistent with the *canonical distance effect* typically observed in the numerical comparison task (Moyer and Landauer, 1967).

The ANOVA performed on the in-order condition of the numerical order verification task showed a significant effect of inter-item distance, $F(1.86, 738.46) = 22.381$, $p < 0.001$, $\eta_p^2 = 0.053$. T-tests revealed that distance 1 trials were significantly faster than distance 2 trials, $t(396) = -5.67$, $p_{FDR\text{-}adj} < 0.001$, $d = -0.285$ and significantly faster than distance 3 trials, $t(396) = -5.28$, $p_{FDR\text{-}adj} < 0.001$, $d = -0.265$. No significant difference was found between distance 2 and distance 3 trials, $t(396) = -0.04$, n.s., $d = 0.002$ (see also **Table 1**). This pattern is consistent with the *reverse distance effect*, i.e., fast reaction times for distance 1 trials compared to distance 2 and 3 trials (Goffin and Ansari, 2016; Vogel et al., 2017). The effect sizes of the *reverse distance effect* are, however, small.

Individual Difference Structure for the Reverse and Canonical Distance Effects

To investigate the structure of individual differences associated with the distance effects, we assessed whether the two effects are best described by the unconstrained or the positive effects model. **Figure 2** shows the results of the modeling for both distance effects, **Table 3** depicts the results of the Bayes factor comparisons between the four models.

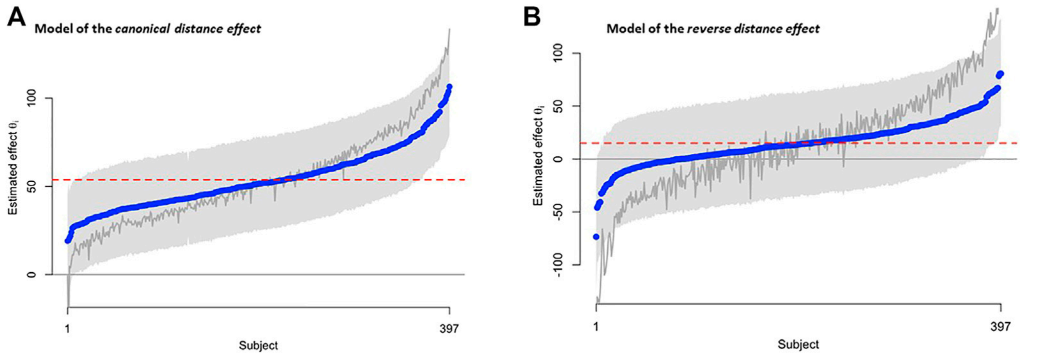


FIGURE 2 | Observed and estimated individual differences in **(A)** the *canonical* and **(B)** the *reverse distance effect* (in ms), ordered by observed effect size. Model-estimated effects (in blue) are derived from the positive-effect model (positive values) for the *canonical distance effect* and from the unconstrained model (positive and negative values) for the *reverse distance effect*. The red dashed line represents the point estimate for the *canonical* and *reverse distance effect respectively*. The gray line represents observed *canonical* and *reverse distance effects*. The gray shaded area denotes the 95% credible interval for the model-estimated effects.

TABLE 3 | Bayes Factor Model Comparisons for **(A)** the canonical distance effect and **(B)** the reverse distance effect.

a) Canonical Distance effect

Prior specification	Unconstrained	Positive-effects	Common-effect	Null
$r_\nu = 1/6, r_\theta = 1/10$	0.19	*	10^{-51}	≈ 0
$r_\nu = 1/12, r_\theta = 1/20$	0.18	*	10^{-51}	≈ 0
$r_\nu = 1/3, r_\theta = 1/5$	0.18	*	10^{-52}	≈ 0

b) Reverse Distance effect

Prior specification	Unconstrained	Positive-effects	Common-effect	Null
$r_\nu = 1/6, r_\theta = 1/10$	*	≈ 0	10^{-13}	10^{-24}
$r_\nu = 1/12, r_\theta = 1/20$	*	≈ 0	10^{-8}	10^{-19}
$r_\nu = 1/3, r_\theta = 1/5$	*	≈ 0	10^{-9}	10^{-19}

Note: The preferred model for each analysis is denoted by an asterisk (*). The remaining cells show the Bayes factor for the indicated model over the preferred model.

For the estimates of the *canonical distance effect*, the unconstrained model and null model received almost no support from the data (see **Table 3A**). Different prior specifications did not change the overall picture. The positive-effects model remained the preferred model across all prior settings (by a factor of approximately 5.6 over the unconstrained model each time), with little support for either the common-effect or null model. Thus, our observed data are evidential for the positive-effects model, indicating quantitative differences in the *canonical distance effect*.

The picture for the estimates of the *reverse distance effect* is quite different, as the unconstrained model was the preferred model. The Bayes factor model comparison (see **Table 3B**) showed that across the three different sets of prior specifications, the unconstrained model was the preferred model by a large factor. The data lend virtually no support for the positive-effects model, common-effect model, nor the null model.

These analyses confirmed that whereas individual differences in the estimates of the *canonical distance effect* appear to be quantitative (i.e., everyone exhibits a positive *canonical distance effect* of some varying magnitude), individual differences in the estimates of the *reverse distance effect* are qualitative (i.e., some exhibit a positive effect, but others show a negative effect). As such we further classified the type of the qualitative distribution of this *distance effect*. This analysis revealed that 172 (43%) participants exhibited a "positive" expected *reverse distance effect*, 18 (5%) individuals showed evidence for a "negative" *reverse distance effect* (i.e., a *canonical distance effect*), and the remaining 207 (52%) were classified as "undecided".

Associations of the Distance Effects With Arithmetic and Mathematical Performance

Our next step was to elucidate the impact of qualitative individual differences of the estimates of the *reverse distance effect* on the association with arithmetic and math performance. We first calculated zero-order and partial correlations for the entire sample (see **Supplementary Material** for the full correlations matrix). While the results revealed significant associations between the estimates of the *canonical distance effect* and all measures of mathematical competence (correlation coefficients range from −0.14 to −0.33; see **Table 4A**), only two significant correlations were found for the estimates of the *reverse distance effect* with small subtractions, $r = -0.11$, $p_{FDR\text{-}adj} < 0.05$, and small additions, $r = -12$, $p_{FDR\text{-}adj} < 0.05$.

TABLE 4 | Bivariate and partial correlations among the *distance effects* and measures of mathematical competence across **(A)** the entire sample, **(B)** the "positive effect" sample and **(C)** the "no-effect" sample.

	1	2	3	4	5	6	7	8
a) Total sample (*n* = 397)								
reverse distance effect	−0.11[a]	−0.12[a]	−0.09	−0.08	−0.02	−0.08	−0.06	−0.03
(partial correlations)	(−0.07)	(−0.09)	(−0.06)	(−0.05)	(0.01)	(−0.06)	(−0.05)	-
canonical distance effect	−0.33[c]	−0.25[c]	−0.21[c]	−0.24[c]	−0.14[b]	−0.17[c]	−0.18[c]	0.05
(partial correlations)	(−0.32[c])	(−0.25[c])	(−0.21[c])	(−0.24[c])	(−0.15[b])	(−0.17[b])	(−0.17[b])	-
b) "positive effect" sample (*n* = 172)								
reverse distance effect	−0.26[c]	−0.28[c]	−0.21[b]	−0.22[b]	−0.09	−0.13	−0.22[b]	−0.09
(partial correlations)	(−0.23[b])	(−0.26[c])	(−0.19[a])	(−0.20[b])	(−0.07)	(−0.12)	(−0.22[b])	-
canonical distance effect	−0.29[c]	−0.23[c]	−0.16	−0.16	−0.12	−0.15	−0.21[b]	0.16
(partial correlations)	(−0.27[b])	(−0.22[a])	(−0.16)	(−0.14)	(−0.11)	(−0.13)	(−0.16)	-
c) "no-positive effect" sample (*n* = 225)								
reverse distance effect	0.12	0.04	0.02	0.09	0.06	0.01	0.04	−0.01
(partial correlations)	(0.13)	(0.04)	(0.02)	(0.09)	(0.06)	(0.01)	(0.04)	-
canonical distance effect	−0.34[c]	−0.26[c]	−0.24[c]	−0.29[c]	−0.15[a]	−0.17[a]	−0.14[a]	−0.05
(partial correlations)	(−0.34[c])	(−0.25[c])	(−0.24[b])	(−0.28[c])	(−0.14)	(−0.16[a])	(−0.16[a])	-

[a] $p_{\text{FDR-adj}} < 0.05$.
[b] $p_{\text{FDR-adj}} < 0.01$.
[c] $p_{\text{FDR-adj}} < 0.001$.
Note: 1, small subtractions; 2, small additions; 3, small multiplications; 4, large subtractions; 5, large additions; 6, large multiplications; 7, M-PA; 8, age; Partial correlations are shown in parenthesis.

Correlation coefficients ranged from −0.2 to −0.12. These two associations were non-significant when the estimates of the *canonical distance effect* and age were included as control variables. In contrast, all correlations of the estimates of the *canonical distance effects* remained significant when considering the estimates of the *reversed distance effect* and age (see also **Table 4A**).

We now investigated these associations in the group of individuals (n = 172) who showed evidence for a "positive" *reverse distance effect* (i.e., individuals who showed a *reverse distance effect*) and those who showed no positive effect ("no-positive effect" group; i.e., undecided and "negative" distance effect; n = 225). Results of the correlation analysis (see **Table 4B**) revealed significant associations between the *reverse distance effect* and small subtractions, small additions, large subtraction, small multiplication, and the M-PA in the "positive" group. Correlation coefficients ranged from −0.09 to −0.28. The above reported correlations remained significant when considering the estimates of the *canonical distance effect* and age in the partial correlation (r values ranged from −0.07 to −0.26). In contrast, the zero order and partial correlation analyses (see **Table 4C**) in the "no-positive effect" group revealed no significant associations with arithmetic operations and the M-PA. Correlation coefficients ranged from 0.01 to 0.12 in the zero-order and from 0.01 to 0.13 in the partial correlation analysis. The performed z-test showed that correlation coefficients in the "positive effect" group compared to the "no-positive effect" group were significantly larger for small subtractions, $z = 3.788$, $p_{FDR\text{-}adj} < 0.05$, small additions, $z = 3.210$, $p_{FDR\text{-}adj} < 0.05$, small multiplications, $z = 2.186$, $p_{FDR\text{-}adj} < 0.01$, large subtraction, $z = 3.075$, $p_{FDR\text{-}adj} < 0.01$, and the M-PA, $z = 2.779$, $p_{FDR\text{-}adj} < 0.01$. No differences were found for large additions, $z = 1.472$, $p_{FDR\text{-}adj}$ < n.s., and large multiplications, $z = 1.379$, $p_{FDR\text{-}adj}$ < n.s.

DISCUSSION

Numerical order processing has been proposed as a significant predictor of arithmetic abilities (for a review see Lyons et al., 2016). However, research on the relationship between the *reverse distance effect* (i.e., as an index of numerical order processing) and arithmetic abilities has demonstrated mixed findings: some studies have found negative (Goffin and Ansari, 2016), positive (Vogel et al., 2019) or no relationship (Vogel et al., 2015; Vogel et al., 2017; Vos et al., 2017; Orrantia et al., 2019). In the present work, we provided evidence that the estimates of the *reverse distance effect* are subject to qualitative individual differences (i.e., not all participants show a *reverse distance effect*) and that these individual differences can obscure the relationship with arithmetic abilities.

We first demonstrated the presence of the *reverse* and the *canonical distance effect* in an ordinal verification and a numerical comparison task. While overall reaction times of the ordinal verification task were faster for small distances (e.g., 2 3 4) compared to large distances (e.g., 2 4 6), reaction times of the numerical comparison task were slower for small distances (e.g., 2 3) compared to large distances (e.g., 8 2). This finding replicates the well-documented behavioral signatures of the *reverse* and *canonical distance effects* in a group of 397 adults. The observed reaction time differences—i.e., fast reaction times for small distances in the ordinal verification task, and slower reaction times for small distances in the numerical comparison task—have been interpreted as evidence for the involvement of different cognitive processing mechanisms (Turconi et al., 2006; Vogel et al., 2015; Lyons et al., 2016). While the *reverse distance effect* has been related to multiple strategies such as long-term memory retrieval and sequential-procedural comparisons (Lyons et al., 2016; Sasanguie and Vos, 2018; Vogel et al., 2019; Sella et al., 2020; Sommerauer et al., 2020), the canonical distance effect has been associated to the mental representation of numerical quantities (Moyer and Landauer, 1967). Independent of the cognitive mechanisms that generate the *reverse* and the *canonical distance effects,* the present data support the view that both *distance effects* are unrelated with one another as we only found a small positive effect size correlation (see full correlation table in **Supplementary Material**) between the two indices ($r_{positive\text{-}effect\ sample} = 0.11$).

The comparison of four hierarchical Bayesian models further showed that individual differences of the estimates of the *reverse distance effect* were best explained by an unconstrained model. Since the unconstrained model allows variation among all possible values (positive or negative), the above results indicate a qualitative structure of the estimates of the reverse distance effect: not all individuals show the expected *reverse distance effect* (Haaf and Rouder, 2017; Haaf and Rouder, 2019). Our estimations revealed that 42% of the individuals demonstrated a *reverse distance effect*, 5% of the individuals showed evidence for an opposite effect (i.e., a *canonical distance effect*), and 52% of the individuals showed no evidence for either direction (i.e., no distance effect). This finding is consistent with the notion that individuals may employ different qualitative processing strategies during the ordinal verification task. Some individuals might use strategies that lead to the *reverse distance effect* (e.g., memory retrieval for small distances in combination with sequential-procedural comparisons for larger distances), while others might use strategies that lead to an opposite effect (e.g., a *canonical distance effect* that arises because sequential-procedural comparisons are used across all distances). An interesting finding is that a large proportion of individuals (52%) showed no evidence for a positive or negative distance effect in the ordinal verification task. In other words, the model was not able to capture whether a distance effect existed in these individuals (i.e., *reverse* or a *canonical distance effect*). We think that there are several possible explanations for this finding. First, it could be the case that our experimental design was not able to detect existing, albeit subtle effects due to a lack of power. However, this is unlikely, as we explicitly tested this possibility with the common-effect and null models. If either of these models had admitted better predictive adequacy, it would call into question our ability to detect individual variations (or any effect at all). Neither model received any support from the data, so we are confident that the issue does not lie within the experimental design. Second, it could be that those individuals showed no distance effect because there is no distance effect to be detected: that is, they have a true effect of zero. Such a scenario could be tested by implementing a mixture modeling approach (e.g., a spike-and-slab model, see also Haaf and Rouder, 2019). Uncovering the reasons of such an absence would be of great interest, since it indicates that these individuals might use a combination of strategies that level each other out (e.g., memory retrieval and sequential comparison that produce opposite effects and zero each other out) or strategies that do not fit with the current models of numerical order processing. For instance, individuals might recognize that the ordinal verification of those triplets, which contain one odd/even and two even/odd numbers (e.g., 2 3 4; 3 6 9), is determined by the position of these numbers in the triplet (e.g., if the odd number is in the middle, it is correct: 2 **3** 4, 3 **6** 9; if it is at the beginning or at end, it is incorrect; 4 2 **3**, **3** 4 2, 3 9 **6, 6** 3 9). Such strategies would be based on non-semantic evaluations and, therefore, shortcut distance related measures. However, the present data are agnostic as to which processing strategies might have been employed. It also leaves unanswered to which extent the observed patterns generalize to other ordinal verification tasks (e.g., with two-digit numbers). These open questions need to be addressed in the future.

In contrast to the ordinal verification task, individual differences of the estimates of *canonical distance effect* were best explained by a positive-effects model. Since the positive-effects model limits individual variations to positive values, the above results indicate quantitative individual differences: all individuals show the expected *canonical distance effect,* but to a different degree. This finding is consistent with neurocognitive models that suggest a continuous/approximate representation of numerical quantities (Moyer and Landauer, 1967). Individual differences may arise as a factor of how much the mental representations of numerical quantities overlap (for a review see Brannon, 2006)—individuals with small representational overlap show less susceptibility and, therefore, a small individual *canonical distance effect*.

We then demonstrated that the association of the *reverse distance effect* with arithmetic abilities is moderated by the observed qualitative individual differences. While neglectable to small correlation coefficients were observed in the no-positive effect sample (i.e., individuals who showed no evidence for a *reverse distance effect*), significant and larger correlations were observed in the positive-effect group (i.e., individuals who showed the expected *reversed distance effect*). This pattern is line with our hypothesis that qualitative differences obscure the relationship between the *reverse distance effect* and arithmetic abilities. Different sample compositions across different studies could, therefore, explain the mixed results that have been reported in previous studies (Goffin and Ansari, 2016; Vogel et al., 2017; Vogel et al., 2019; Vos et al., 2017; Orrantia et al., 2019). This finding highlights the need to pay close attention to the sample composition and to ensure that the dimension of interest (e.g., the *reverse distance effect*) is not confounded by qualitative processing differences. When controlling for this confound, we observed significant negative associations between both the *reverse distance effects* and arithmetic abilities as well as between the *canonical distance effect* and arithmetic abilities. Thus, the results of this analysis are in line with the findings reported by Goffin and Ansari (2016) who reported negative correlations with arithmetic fluency measures of the Woodcock-Johnson III Test of Achievement (Woodcock et al., 2001) for both distance effects ($r = -0.422$ for the *reverse distance effect* and r = −0.310 for the *canonical distance effect*). Together, these findings indicate that individuals with smaller distance effects (individuals who are less susceptible to the influence of numerical distances) show better arithmetic performances. The question of which cognitive mechanisms give rise to these associations needs to be further explored in future studies.

Previous studies that have investigated the association between the *reverse distance effect* and arithmetic have predominantly used composite scores (i.e., a combination of different arithmetic performance measures) to index arithmetic abilities. This approach neglects that arithmetic consists of different operations (e.g., subtractions, additions, multiplications, divisions) and that different strategies are used to solve them (see Ashcraft, 1992; Campbell and Xue, 2001). For instance, there is good neurocognitive evidence that individuals use different strategies and procedures (such as fact-retrieval or calculation) to find the correct answer to arithmetic problems (e.g., Grabner et al., 2009). The results of our study indicated significant associations of the *reverse distance effect* with all small problems (i.e., subtraction, addition and multiplication) and large subtraction problems. No significant associations were found for large additions and large multiplications. For the *canonical distance effect*, significant associations with small subtraction and additions were found. All other arithmetic operations did not significantly correlate with the *canonical distance effect*.

From a theoretical standpoint, one might have expected stronger relationships for the canonical distance effect with arithmetic operations that require the manipulation of numerical quantities, i.e., large problems instead of small problems, which are often solved via fact retrieval (LeFevre et al., 1996; Campbell and Xue, 2001). The present results contradict this in as much only small subtraction and small additions were found to correlate with the canonical distance effect. However, some findings suggest that small additions and small subtractions can be solved via fast and automatic procedures of numerical processing (for a discussion on small addition see Baroody, 2018). This is in contrast to small multiplications, which have been argued to be the prime example of fact-retrieval (Ashcraft, 1992; Campbell and Epp, 2005). As such it is possible that the present results capture that difference within the small problem range. It is also possible that the larger problems in our paper-pencil task, which were quite complex and in which individuals might have used different strategies and other cognitive processing mechanisms to find the solution (e.g., in the large problem task individuals had to carry over the results), did not capture the manipulation of numerical quantities. For the *reverse distance effect,* one might have expected less specific and rather broad association as the *reverse distance effect* is argued to arise from a combination of different strategies. One could, however, argue that the fast-retrieval of ordinal relationships drives the *reverse distance effect* (Vogel et al., 2019; Sella et al., 2020), and that associations with arithmetic problems that afford the fast access to stored knowledge (i.e., small problems) are to be expected. To some extent that could explain the stronger relationship of the *reverse distance effect* to all small problems. Taken together, although these explanations are speculative, the data suggest distinctive associations of the two distance effects with different facets of arithmetic operations.

We also found a significant relationship between the *reverse distance effect* and more complex forms of mathematics (i.e., the M-PA). The association between numerical order processing and complex forms of mathematics has, to the best of our knowledge, only been investigated in two other studies. Morsanyi et al. (2018) collected data from 87 undergraduate students who performed a numerical order task, a number-line task, a test of arithmetic abilities (i.e., the math fluency subset of the Woodcock-Johnson III Test of Achievement, Woodcock et al., 2001) and a questionnaire that assesses individual differences in cognitive thinking styles (i.e., preference for object-spatial imagery or verbal cognitive style). The authors found a significant association of numerical order processing with the number line task as well as the self-reported object-spatial thinking style. However, the findings of this study were not based on the *reverse distance effect* as a measure of numerical order processing. The authors rather used less specific composite scores of overall reaction time measures and accuracy rates. Orrantia et al. (2019) investigated the relationship between numerical order processing, arithmetic abilities, and general mathematical achievement (i.e., the Spanish adaption of the SRA Test of Educational Ability) in a group of 27 male university students. The results of this study did not find a significant association between these measures. Despite the small sample size for a correlational investigation, the authors did also not use the *reverse distance effect* to investigate a possible association of these measures. Thus, the findings of the present work extend these studies by suggesting a specific association between the *reverse distance effect* and more complex forms of mathematical reasoning.

We conclude that the present work provided evidence for qualitative individual differences of the *reverse distance effect* (i.e., not all participants show this expected effect) and that this individual variation can obscure the relationship with arithmetic abilities and other mathematical competencies (e.g., depending on the sample composition and the individuals that show an effect). When controlling for these individual differences, we found a significant relationship between variations of the *reverse distance effect* and different measures of arithmetic and mathematical performance. The reasons for the observed qualitative differences in the numerical order verification task remain, however, unclear and need to be further investigated. To achieve this, future work needs to ensure that dimensions of interest (i.e., here the *reverse distance effect*) are not confounded by qualitative differences.

DATA AVAILABILITY STATEMENT

The dataset presented in this study can be accessed via the open science framework (OSF) using the following link: https://osf.io/jvmc2/?view_only=64abdcbd0cac4ca0b8154ebbcc4437a2.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

SV contributed to conceptualizing the idea, analyzing the data, writing the first complete draft, and editing the paper. TF contributed to conceptualizing the idea, analyzing the data, and editing the paper. RG contributed to the editing of the paper.

ACKNOWLEDGMENTS

We like to thank Dennis Wambacher for the implementation of the task and our research assistants for collecting this data set. The authors acknowledge the financial support by the University of Graz.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2021.655747/full#supplementary-material

REFERENCES

Ashcraft, M. H. (1992). Cognitive Arithmetic: A Review of Data and Theory. *Cognition* 44 (1), 75–106. doi:10.1016/0010-0277(92)90051-I

Ashkenazi, S., Henik, A., Ifergane, G., and Shelef, I. (2008). Basic Numerical Processing in Left Intraparietal Sulcus (IPS) Acalculia. *Cortex* 44 (4), 439–448. doi:10.1016/j.cortex.2007.08.008

Baroody, A. J. (2018). A Commentary on Chen and Campbell (2017): Is There a clear Case for Addition Fact Recall? *Psychon. Bull. Rev.* 25 (6), 2398–2405. doi:10.3758/s13423-018-1440-y

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Brannon, E. M. (2006). The Representation of Numerical Magnitude. *Curr. Opin. Neurobiol.* 16 (2), 222–229. doi:10.1016/j.conb.2006.03.002

Campbell, J. I. D., and Epp, L. J. (2005). "Architectures for Arithmetic," in *Handbook of Mathematical Cognition*. Editors J. I. D. Campbell, (Psychology Press), New York: NY 347–360. doi:10.4324/9780203998045

Campbell, J. I. D., and Xue, Q. (2001). Cognitive Arithmetic across Cultures. *J. Exp. Psychol. GeneralGeneral* 130 (2), 299–315. doi:10.1037/0096-3445.130.2.299

Cohen, J., Cohen, P., West, S. G., , et al.Cohen, P., West, S. G., Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. New York: Routledge, 536. doi:10.4324/9780203774441

De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How Do Symbolic and Non-symbolic Numerical Magnitude Processing Skills Relate to Individual Differences in Children's Mathematical Skills? A Review of Evidence from Brain and Behavior. *Trends Neurosci. Edu.* 2 (2), 48–55. doi:10.1016/j.tine.2013.06.001

De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The Predictive Value of Numerical Magnitude Comparison for Individual Differences in Mathematics Achievement. *J. Exp. child Psychol.* 103 (4), 469–479. doi:10.1016/j.jecp.2009.01.010

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* 88 (1), 2–19. doi:10.1016/j.neuron.2015.09.019

Delazer, M., Karner, E., Zamarian, L., Donnemiller, E., and Benke, T. (2006). Number Processing in Posterior Cortical Atrophy-A Neuropsycholgical Case Study. *Neuropsychologia* 44 (1), 36–51. doi:10.1016/j.neuropsychologia.2005.04.013

Faulkenberry, T. J., and Bowman, K. (2020). *Modeling the Latent Structure of Individual Differences in the Numerical Size-Congruity Effect* New York, NY: PsyArXiv. doi:10.31234/osf.io/4b9rs

Faulkenberry, T. J., Ly, A., and Wagenmakers, E.-J. (2020). Bayesian Inference in Numerical Cognition: A Tutorial Using JASP. *J. Numer. Cogn.* 6 (2), 231–259. doi:10.5964/jnc.v6i2.288

Field, A., Miles, J., and Field, Zoe. (2012). *Discovering Statistics Using R* London, England: SAGE Publication Ltd. https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-r/book236067.

Franklin, M. S., Jonides, J., and Smith, E. E. (2009). Processing of Order Information for Numbers and Months. *Mem. Cogn.* 37 (5), 644–654. doi:10.3758/MC.37.5.644

French, J. W., Ekstrom, R. B., and Price, I. A. (1963). *Kit for Reference Tests for Cognitive Factors* Princeton, NJ: Educational Testing Service. doi:10.21236/ad0410915

Goffin, C., and Ansari, D. (2016). Beyond Magnitude: Judging Ordinality of Symbolic Number Is Unrelated to Magnitude Comparison and Independently Relates to Individual Differences in Arithmetic. *Cognition* 150, 68–76. doi:10.1016/j.cognition.2016.01.018

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., and Neuper, C. (2009). To Retrieve or to Calculate? Left Angular Gyrus Mediates the Retrieval of Arithmetic Facts during Problem Solving. *Neuropsychologia* 47 (2), 604–608. doi:10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., and De Smedt, B. (2011). Neurophysiological Evidence for the Validity of Verbal Strategy Reports in Mental Arithmetic. *Biol. Psychol.* 87 (1), 128–136. doi:10.1016/j.biopsycho.2011.02.019

Greenhouse, S. W., and Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika* 24 (2), 95–112. doi:10.1007/BF02289823

Gross, J., Hudson, C., and Price, D. (2009). *The Longterm Costs of Numeracy Difficulties* London, England: Every Child a Chance Trust and KPMG.

Haaf, J. M., and Rouder, J. N. (2017). Developing Constraint in Bayesian Mixed Models. *Psychol. Methods* 22 (4), 779–798. doi:10.1037/met0000156

Haaf, J. M., and Rouder, J. N. (2019). Some Do and Some Don't? Accounting for Variability of Individual Difference Structures. *Psychon. Bull. Rev.* 26 (3), 772–789. doi:10.3758/s13423-018-1522-x

Holloway, I. D., and Ansari, D. (2009). Mapping Numerical Magnitudes onto Symbols: The Numerical Distance Effect and Individual Differences in Children's Mathematics Achievement. *J. Exp. Child Psychol.* 103 (1), 17–29. doi:10.1016/j.jecp.2008.04.001

Jasper, F., and Wagener, D. (2011). *Mathematiktest für die Personalauswahl [Mathematics test for selection of personnel]*. Göttingen: Hogrefe.

Jeffreys, S. H. (1968). *The Theory of Probability*. Third Edition New York, NY: Oxford University Press.

Kass, R. E., and Raftery, A. E. (1995). Bayes Factors. *J. Am. Stat. Assoc.* 90 (430), 773–795. doi:10.1080/01621459.1995.10476572

LeFevre, J.-A., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., and Sadesky, G. S. (1996). Multiple Routes to Solution of Single-Digit Multiplication Problems. *J. Exp. Psychol. Gen.* 125 (3), 284–306. doi:10.1037/0096-3445.125.3.284

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2011). Symbolic and Non-symbolic Distance Effects in Children and Their Connection with Arithmetic Skills. *J. Neurolinguist.* 24 (5), 583–591. doi:10.1016/j.jneuroling.2011.02.004

Lyons, I. M., and Ansari, D. (2015). Numerical Order Processing in Children: From Reversing the Distance-Effect to Predicting Arithmetic. *Mind, Brain Edu.* 9 (4), 207–221. doi:10.1111/mbe.12094

Lyons, I. M., and Beilock, S. L. (2011). Numerical Ordering Ability Mediates the Relation between Number-Sense and Arithmetic Competence. *Cognition* 121 (2), 256–261. doi:10.1016/j.cognition.2011.07.009

Lyons, I. M., and Beilock, S. L. (2013). Ordinality and the Nature of Symbolic Numbers. *J. Neurosci.* 33 (43), 17052–17061. doi:10.1523/JNEUROSCI.1775-13.2013

Lyons, I. M., Vogel, S. E., and Ansari, D. (2016). On the Ordinality of Numbers. *Prog. Brain Res.* 227, 187–221. doi:10.1016/S1099-4831(10)06803-310.1016/bs.pbr.2016.04.010

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., and Fugelsang, J. (2010). Challenging the Reliability and Validity of Cognitive Measures: The Case of the Numerical Distance Effect. *Acta Psychologica* 134 (2), 154–161. doi:10.1016/j.actpsy.2010.01.006

Moyer, R. S., and Landauer, T. K. (1967). Time Required for Judgements of Numerical Inequality. *Nature* 215 (2), 1519–1520. doi:10.1038/2151519a0

Nuerk, H.-C., Weger, U., and Willmes, K. (2001). Decade Breaks in the Mental Number Line? Putting the Tens and Units Back in Different Bins. *Cognition* 82 (1), B25–B33. doi:10.1016/S0010-0277(01)00142-1

Orrantia, J., Muñez, D., Matilla, L., Sanchez, R., San Romualdo, S., and Verschaffel, L. (2019). Disentangling the Mechanisms of Symbolic Number Processing in Adults' Mathematics and Arithmetic Achievement. *Cogn. Sci.* 43 (1). doi:10.1111/cogs.12711

Parsons, S., and Bynner, J. (2005). *Does Numeracy Matter More.* London: NRDC (National Research and Development Centre for adult literacy and numeracy). [aRCK].

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Front. Neuroinform.* 2, 10. doi:10.3389/neuro.11.010.2008

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., and Ansari, D. (2007). Impaired Parietal Magnitude Processing in Developmental Dyscalculia. *Curr. Biol.* 17 (24), R1042–R1043. doi:10.1016/j.cub.2007.10.013

R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Revelle, W. (2020). Psych: Procedures for Personality and Psychological Research. Version 2.0.12. [Computer software]. Available at: https://cran.r-project.org/web/packages/psych/psych.pdf.

Rousselle, L., and Noël, M.-P. (2007). Basic Numerical Skills in Children with Mathematics Learning Disabilities: A Comparison of Symbolic vs Non-symbolic Number Magnitude Processing. *Cognition* 102 (3), 361–395. doi:10.1016/j.cognition.2006.01.005

Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012). Association between Basic Numerical Abilities and Mathematics Achievement. *Br. J. Dev. Psychol.* 30 (2), 344–357. doi:10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., and Vos, H. (2018). About Why There Is a Shift from Cardinal to Ordinal Processing in the Association with Arithmetic between First and Second Grade. *Dev. Sci.* 21 (5), e12653. doi:10.1111/desc.12653

Schillinger, F. L., Vogel, S. E., Diedrich, J., and Grabner, R. H. (2018). Math Anxiety, Intelligence, and Performance in Mathematics: Insights from the German Adaptation of the Abbreviated Math Anxiety Scale (AMAS-G). *Learn. Individual Differences* 61, 109–119. doi:10.1016/j.lindif.2017.11.014

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of Non-symbolic and Symbolic Numerical Magnitude Processing with Mathematical Competence: A Meta-Analysis. *Dev. Sci.* 20 (3), e12372. doi:10.1111/desc.12372

Schnuerch, M., Nadarevic, L., and Rouder, J. N. (2020). The Truth Revisited: Bayesian Analysis of Individual Differences in the Truth Effect. *Psychon. Bull. Rev.* 28, 750–765. doi:10.3758/s13423-020-01814-8

Sella, F., Sasanguie, D., and Reynvoet, B. (2020). Judging the Order of Numbers Relies on Familiarity rather Than Activating the Mental Number Line. *Acta Psychologica* 204, 103014. doi:10.1016/j.actpsy.2020.103014

Sommerauer, G., Graß, K.-H., Grabner, R. H., and Vogel, S. E. (2020). The Semantic Control Network Mediates the Relationship between Symbolic Numerical Order Processing and Arithmetic Performance in Children. *Neuropsychologia* 141, 107405. doi:10.1016/j.neuropsychologia.2020.107405

Turconi, E., Campbell, J. I. D., and Seron, X. (2006). Numerical Order and Quantity Processing in Number Comparison. *Cognition* 98, 273–285. doi:10.1016/j.cognition.2004.12.002

Van Opstal, F., Gevers, W., De Moor, W., and Verguts, T. (2008). Dissecting the Symbolic Distance Effect: Comparison and Priming Effects in Numerical and Nonnumerical Orders. *Psychon. Bull. Rev.* 15 (2), 419–425. doi:10.3758/PBR.15.2.419

Vogel, S. E., Haigh, T., Sommerauer, G., Spindler, M., Brunner, C., Lyons, I. M., et al. (2017). Processing the Order of Symbolic Numbers: A Reliable and Unique Predictor of Arithmetic Fluency. *J. Numer. Cogn.* 3 (2), 288–308. doi:10.5964/jnc.v3i2.55

Vogel, S. E., Koren, N., Falb, S., Haselwander, M., Spradley, A., Schadenbauer, P., et al. (2019). Automatic and Intentional Processing of Numerical Order and its Relationship to Arithmetic Performance. *Acta Psychologica* 193, 30–41. doi:10.1016/j.actpsy.2018.12.001

Vogel, S. E., Remark, A., and Ansari, D. (2015). Differential Processing of Symbolic Numerical Magnitude and Order in First-Grade Children. *J. Exp. Child Psychol.* 129, 26–39. doi:10.1016/j.jecp.2014.07.010

Vos, H., Sasanguie, D., Gevers, W., and Reynvoet, B. (2017). The Role of General and Number-specific Order Processing in Adults' Arithmetic Performance. *J. Cogn. Psychol.* 29 (4), 469–482. doi:10.1080/20445911.2017.1282490

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Test of Achievement* Itasaca, IL: Riverside.

Zorzi, M., and Butterworth, B. (1999). "A Computational Model of Number Comparison," in *Twenty First Annual Meeting of the*. Editors M. Hahn and S. C. Stones (Vancouver, BC: Cognitive Science Society).

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fpsyg.2021.583734

Check for updates

Development of Preschoolers' Understanding of Zero

Attila Krajcsi[1]*, *Petia Kojouharova*[2,3] *and Gábor Lengyel*[4]

[1] *Cognitive Psychology Department, Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary,* [2] *Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary,* [3] *Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary,* [4] *Department of Cognitive Science, Central European University, Budapest, Hungary*

While knowledge on the development of understanding positive integers is rapidly growing, the development of understanding zero remains not well-understood. Here, we test several components of preschoolers' understanding of zero: Whether they can use empty sets in numerical tasks (as measured with comparison, addition, and subtraction tasks); whether they can use empty sets soon after they understand the cardinality principle (cardinality-principle knowledge is measured with the give-N task); whether they know what the word "zero" refers to (tested in all tasks in this study); and whether they categorize zero as a number (as measured with the smallest-number and is-it-a-number tasks). The results show that preschoolers can handle empty sets in numerical tasks as soon as they can handle positive numbers and as soon as, or even earlier than, they understand the cardinality principle. Some also know that these sets are labeled as "zero." However, preschoolers are unsure whether zero is a number. These results identify three components of knowledge about zero: operational knowledge, linguistic knowledge, and meta-knowledge. To account for these results, we propose that preschoolers may understand numbers as the properties of items or objects in a set. In this view, zero is not regarded as a number because an empty set does not include any items, and missing items cannot have any properties, therefore, they cannot have the number property either. This model can explain why zero is handled correctly in numerical tasks even though it is not regarded as a number.

Keywords: numerical cognition, zero, number status of zero, items based number representation, cardinality principle

OPEN ACCESS

Edited by:
Catherine Sandhofer,
University of California, Los Angeles, United States

Reviewed by:
Ren Liu,
University of California, Merced, United States
Larissa Salvador,
Universidade Federal de Minas Gerais, Brazil

Correspondence:
Attila Krajcsi
krajcsi@gmail.com

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

Received: *18 September 2020*
Accepted: *05 July 2021*
Published: *27 July 2021*

Citation:
Krajcsi A, Kojouharova P and Lengyel G (2021) Development of Preschoolers' Understanding of Zero. Front. Psychol. 12:583734. doi: 10.3389/fpsyg.2021.583734

HIGHLIGHTS

- Preschoolers can handle zero as soon as they can handle positive integers.
- Preschoolers are unsure whether zero is a number.
- Children may start to understand numbers as the properties of items in a set.

INTRODUCTION

Children start to understand the use of symbolic exact numbers at around the age of three (Wynn, 1990, 1992). Although many details on the development of understanding natural numbers are already known, the development of understanding zero remains mostly unknown, and it is not integrated into any numerical cognition models. It is still largely unknown how zero is handled and understood in tasks in which symbolic natural numbers are already used successfully by preschoolers. The main aim of the present study is to describe more fully the development of understanding zero and to consider its theoretical implications.

Lack of Developmental Models for Understanding Zero

Models concerning the development of numerical cognition mostly cannot specify how understanding zero is integrated into more general numerical knowledge. As a starting point, in the infant literature, there is agreement that non-symbolic numerical information (such as arrays of dots, and series of sounds or events) is probably processed by two representations (Feigenson et al., 2004; Piazza, 2010). Dominant models propose that, in infants, numerical information is handled by either the imprecise Approximate Number System (Feigenson et al., 2004; Piazza, 2010) or the visual attention related Object Tracking System (Feigenson et al., 2004). However, it is not straightforward whether either of these systems can handle zero (see more details on how these models may or may not account for zero processing in **Supplementary Material**).

The next important step in the development of number understanding is for preschoolers to acquire an understanding of exact large symbolic numbers. Symbolic numbers are values that are denoted by symbols (in the case of preschoolers, such symbols are usually number words), as opposed to non-symbolic quantities, such as arrays of visual objects or series of auditory events. According to the consensus in the literature, at around the age of three or four, children start to understand the conceptual principles of number use, which is referred to as understanding the cardinality principle (Wynn, 1990; Lipton and Spelke, 2006; Sarnecka and Carey, 2008). With this principle, they are able to handle exact symbolic numbers. Number knowledge in preschoolers is usually measured with the "give a number" (or give-N) task, in which children are asked to give a specific number of objects from a pile of objects (Wynn, 1990, 1992). With this task, one can determine what phase of number understanding a child is in. The first phase is the pre-numeric phase; in this phase, although preschoolers know the counting list (i.e., the series of number words starting with "one-two-three"), they do not know the meaning of these words and fail in the task. These children are termed pre-knowers. The second phase is when children become subset-knowers; they can give 1, 2, 3, or 4 items, but not when asked for more, even when they know how the counting list continues. The final phase is when preschoolers become cardinality-principle-knowers (CP-knowers); they can give any amount of items that is in their known counting list. This phase is believed to show their real understanding of exact symbolic numbers (Wynn, 1990, 1992; Lipton and Spelke, 2006; Sarnecka and Carey, 2008). (Although see some limitations of this description, for example, in Le Corre and Carey, 2007; Davidson et al., 2012; Le Corre, 2014; Sella and Lucangeli, 2020).

There is no consensus on what representational changes occur for a preschooler to understand the cardinality principle. Most models suppose that, in some way, the systems available in infancy may play a role in the first steps (Carey, 2004, 2009; Piazza, 2010), although it is not known entirely how these or other systems contribute to their understanding of symbolic exact numbers. Relatedly, it is not known whether preschoolers understand symbolic zero when they understand the cardinality principle. Importantly, the few works that have investigated preschoolers' symbolic understanding of zero (see the following subsection) cannot be integrated into this framework because those works did not investigate whether or not the preschoolers understood the cardinality principle.

Contradictory Results on Preschoolers' Understanding of Zero

To our knowledge, there are only two studies on preschoolers investigating quantitatively the zero concept when it is denoted symbolically (Wellman and Miller, 1986; Bialystok and Codd, 2000). Note again that the present work focuses on the processing of symbolic stimuli because (a) investigating non-symbolic zero involves many unresolved methodological issues, (b) in recent years, several works have revealed essential differences between symbolic and non-symbolic number processing (e.g., Noël and Rousselle, 2011; Bulthé et al., 2015; Krajcsi et al., 2016, 2020; Schneider et al., 2017), and such differences put into question whether symbolic stimuli are processed by an evolutionarily old, imprecise number representation, and (c) CP-knowledge points beyond approximate number handling (Carey and Barner, 2019)[1]. As mentioned above, neither of these two studies is related to the current developmental models. Importantly, the

[1]There are a few other related former works that are not relevant from the viewpoint of the present work. First, some of the works cited in the main text also investigate children in their first school years. Still, we mostly focus on the preschool years, as our study investigates knowledge of zero in preschoolers around the time when the cardinality principle is acquired. Second, while several former works investigated the understanding of zero in preschoolers, they study other aspects of this understanding not in line with the aims of the present study, and their results are not conclusive for the present aims. While Merritt and Brannon (2013) collect preschooler data about processing zero, the comparison task is not symbolic. Also, while Davidson (1992) uses numerical tasks with zero in preschoolers, those tasks alternate training tasks and questions, so it is not clear what is the effect of the training in the session. Finally, while Baroody et al. (2009) partly investigate symbolic arithmetical operations with zero, they do not report the results of positive number-only tasks; therefore, it is impossible to tell whether the children had more difficulty with zero than with positive numbers, and they measure the number knowledge of the children with the Is it N task, which is considered to be an invalid measure of the number knowledge by the literature (Wynn, 1990, 1992).

Another seemingly related topic is the handling of zero in transcoding tasks. In transcoding tasks, one has to translate a number from a notation to another notation, e.g., read aloud an Indo-Arabic number, where Indo-Arabic notation is transcoded into number word notation. It has been demonstrated that zero digits are often transcoded erroneously, e.g., "two hundred and two" is transcoded into 2002 (Grana et al., 2003; Zuber et al., 2009; Moeller et al., 2015). While this phenomena is related to the present topic in the sense that it is related to zero, there are at least three critical differences. First, while the present work discusses the zero *number* (i.e., the value between 1 and −1), transcoding considers the zero *digit* (i.e., the symbol that is used to denote a missing power in multi-power notation; e.g., the number 10 is a number between 9 and 11, and the zero digit is used only to denote that in a decimal system that a number does not include ones). Second, place-value notational systems, such as the Indo-Arabic number system, are relatively difficult to understand both for adults and children (Krajcsi and Szabó, 2012), and transcoding tasks require not only an understanding of zero, but also an understanding of place-value and other multi-power notational systems. Consequently, it is not straightforward how strongly transcoding issues are rooted in zero digits handling or in multi-power notation processing. Third, syntactic processes behind transcoding are only relevant in multi-power (e.g., multi-digit conclusions of these two works contradict each other on whether or not handling zero is more difficult for preschoolers compared to handling positive integers. In this subsection, we briefly summarize the two studies, and then discuss the potential causes of their differing conclusions.

Wellman and Miller (1986) presented children between 3 and 7 years old with the following tasks: (a) count items in sets (including empty sets), (b) name the smallest number they know, (c) name Indo-Arabic symbols, and (d) compare numbers between 0 and 5 in Indo-Arabic notation. Their main finding was that the children's understanding and use of zero were delayed compared to their use of positive numbers. In their detailed analysis, the authors concluded that there are three typical behavioral patterns or, in their terminology, phases. In the first phase, children can name the 0 symbol, although they do not understand its meaning. In the second phase, children can count backward to zero, with the understanding that zero means nothing. Finally, in the third phase, children know that zero is the smallest number, and they can compare numbers even if one of the numbers is zero. In this description, the progress of their development is slow: 4-year-old children are not yet at the first phase, and only 6-year-old children are at the final phase.

However, a later study found evidence that understanding zero may be as uncomplicated as positive numbers for preschoolers. Bialystok and Codd (2000) investigated preschoolers' understanding and spontaneous notation of positive integers, zero, and fractions. They found that, for preschoolers, understanding zero is not harder than understanding positive integers; this conclusion is not in line with the previously described study by Wellman and Miller (1986). In this work, children between 3 and 7 years old were asked to give different amounts of cookies to puppets and to make written notes about these amounts. The children had to recall the amounts 20 min later; and again 2 weeks later. In both instances, the children were allowed to use the notes they had made. According to the results, the children were able to solve the give-zero task. However, it is important to note that the instruction was not formed in the usual mathematical way, e.g., the children were not told to "give Big Bird zero cookies"; instead they were instructed to "give Big Bird no cookies for lunch." The children were able to make a note of the number zero as efficiently as making a note of positive integers. Similarly, they could recall the correct number after 20 min, be it zero or a positive integer. However, 2 weeks later, unlike 5-year-old children, 3- and 4-year-old children were unable to recall the number zero as successfully as recalling positive integers. (The same result was found by Hughes, 1986, who found that children can use notes for denoting zero, however, the quantitative results of that study were not published).

The contradiction between the conclusions of these two studies (i.e., Wellman and Miller, 1986; Bialystok and Codd, 2000) may originate from methodological differences and from interpretational issues. Obviously, the two studies used completely different tasks, and it is possible that zero can be handled more easily in some tasks than in others. Yet, there are some less trivial sources of differences. First, while Wellman and Miller (1986) suggest three phases of development for understanding zero, their data actually seem to reveal four phases. (Even though the authors mentioned that the data may include some inconsistencies, they nonetheless insisted on an interpretation with three phases.) The additional phase is between the second and third phases: After successfully counting back to zero, the preschoolers could compare numbers with zero, even though they did not know that zero is the smallest number (see Table 1 in Wellman and Miller, 1986). In fact, this phase is paradoxical: While children know that zero is smaller than one, they think that one is the smallest number. What causes this dissociation in their zero-knowledge? As a possible explanation, we hypothesize that children do not think that zero is a number. This possible misconception is even observable in adults: In a study, 15% of preservice elementary school teachers responded that zero is not a number (Wheeler and Feghali, 1983). Another possible explanation is that zero is not part of the counting list (which usually starts with "one"), and this is why children handle zero differently (Merritt and Brannon, 2013). Both explanations suggest that their meta-knowledge of the number status of zero may be independent of handling the zero value correctly. Consequently, one may assume that children can understand zero sufficiently when they compare zero correctly, but they do not yet understand that zero should be categorized as a number. If this is the case, children may understand zero earlier than what was proposed by Wellman and Miller. We return to this problem and to a more detailed list of possible explanations in the discussion section.

A second methodological problem that could be the cause of the two studies' different conclusions is that the linguistic formulation of the tasks including zero could have influenced performance. While the mathematical viewpoint suggests that zero is a number just like any other integer and that zero should therefore be used linguistically in the same way as other numbers, natural language mostly uses different linguistic forms for statements about zero. For example, we usually do not say "The car is traveling with zero kilometers per hour"; instead, we say "The car has stopped (or is stationary)." Similarly, we do not say "Give zero cookies to Peppa Pig"; instead, we say "Do not give any cookies to Peppa Pig" or "Give no cookies to Peppa Pig." Thus, it is possible that using mathematical language is harder for children than using natural language because the former is less familiar to them. This can be hypothetically confirmed by the data from the two studies: The Wellman and Miller (1986) used mathematical language and found that children experienced difficulties in understanding zero, while Bialystok and Codd (2000) used natural language and found that children experienced no difficulties in handling zero. However, based only on these data, one cannot be sure whether language form significantly influenced performance, because of the many other differences between the two studies.

Aims of the Study

To create appropriate models, it is essential to first have reliable data. Considering that the two previously discussed studies

Indo-Arabic) numbers that are typically not understood by preschoolers, but only by older children (Zuber et al., 2009; Moeller et al., 2015).

came to contradictory conclusions on whether processing the zero value compared to processing positive values is harder for preschoolers (Wellman and Miller, 1986; Bialystok and Codd, 2000), and that we cannot confirm the causes of the contradictions relying on the two studies' data, the present study aims to provide additional more systematically collected data on preschoolers' symbolic understanding of zero. Note that this means that our aims do not build upon any theoretical models, such as the Approximate Number System, or the Object Tracking System (presented in the first part of the Introduction), rather, this study seeks to clarify the main phenomena and describe the development more precisely and extensively than previous works have (presented in the second part of the Introduction).

(Aim 1) Specifically, to use a more comprehensive range of tasks (i.e., giving a set, comparison, addition and subtraction) in order to investigate whether children can handle zero in numerical tasks as efficiently as they can handle positive integers (Aim 2). To test the potential effect of language form (contrasting mathematical vs. natural language and investigating whether the number word "zero" is understood) on performance in the tasks (Aim 3). To put these findings into the context of current models for understanding the cardinality principle, the present study investigates whether subset-knowers and cardinality-principle-knowers (as measured with the give-N task) have a different level of understanding of zero (if understanding zero is available at such an early age) (Aim 4). To investigate whether the additional phase we emphasized in the data from Wellman and Miller (1986) is reliably observable, i.e., whether at some point children can compare zero correctly, even though they still think that zero is not a number.

MATERIALS AND METHODS

Participants

Forty 3- and 4-year-old Hungarian preschoolers participated in this study. Because of the methodological limitation of the applied give-N task, the number knowledge of two of these children (i.e., whether they were one-knowers or pre-knowers) could not be specified; consequently, their data were excluded from the study (find more details in the tasks and results sections). The data of 19 boys and 19 girls were analyzed, with a mean age of 4 years and 1 months, ranging between 3 years 2 months and 5 years 1 month. Two preschools were involved in the data collection; one of them in the capital and the other one in a country town (18 preschoolers from the capital, 7 girls in the capital and 12 girls in the country town). The children in both preschools were mostly from middle-class families. None of the preschoolers had previously received formal training on handling zero in preschools. No further data were collected on their sociodemographic characteristics or former numerical training.

Tasks

The tasks used in this study covered three main areas (**Table 1**). Note that the present tasks mainly investigate the handling of symbolic numbers. (1) The give-N task categorized children in terms of whether they were cardinality-principle-knowers (CP-knowers) or subset-knowers. (2) This task was also used with "zero" and "nothing" to determine whether the preschoolers can apply the zero value in the task. Additionally, comparison, addition, and subtraction tasks were used to measure whether the preschoolers could use zero in operations as efficiently as they could use positive integers. These tasks were also used to measure the effect of various linguistic versions of the same tasks. (3) The preschoolers' meta-knowledge of numbers was measured to investigate whether they understood that zero is a number.

TABLE 1 | Summary of the tasks.

Name of the task	Short description
Measuring number knowledge	
Give-N (positive numbers)	Give N balls to an agent.
Operations with zero	
Give-N (nothing and zero)	Give N balls to an agent.
Comparison	Choose the larger set.
Addition	Add two values.
Subtraction	Subtract one value from another.
Meta-knowledge of zero	
Smallest number	Name the smallest number.
Is it a number?	Say whether something is a number or not.

Tasks Related to the Aims

Because the present study includes four independent aims (independent in the sense that any of the aims would be meaningful without the others), and because many of the tasks contribute to several of the aims at the same time, we give explicit guidance throughout the text on what aspects of the tasks or task-combinations contribute to which aims. Here, we review the aims and what aspects of the tasks investigate those aims. Aim 1 (is zero more difficult to handle than positive integers for preschoolers) is measured with the operations tasks. The relevant contrast is whether zero-related operations compared to positive number-related operations show worse performance. Aim 2 (the role of the linguistic form in understanding zero) is measured across all tasks: The give-N task measures the preschoolers' interpretation of the "zero" and "nothing" labels in that task; the operations tasks measure the effect of the mathematical and natural linguistic forms; and the meta-knowledge tasks investigate again the "zero" and "nothing" labels in those contexts. The relevant contrast is whether different linguistic versions induce different levels of performance. Aim 3 (the role of number knowledge in terms of subset-knowers and CP-knowers) is measured in both the operations and meta-knowledge tasks by contrasting the two number-knowledge groups. Finally, Aim 4 (do preschoolers lack meta-knowledge about zero when they can handle zero in operations) is investigated in the meta-knowledge tasks, whose results are contrasted with the results of the operations tasks (See also the analysis plan below).

Give-N Task

In this task a pile of balls was in view of the children, and they had to give a specific number of balls to a toy bird. The task measures

(a) whether the child understands the cardinality principle, and (b) whether the child understands what the words "nothing" and/or "zero" refer to.

Note that while the performance with positive numbers has been investigated and validated extensively (for example see the seminal works of Wynn, 1990, 1992), to our knowledge, performance with zero in this task has not been studied yet. Also, to our knowledge, understanding zero and understanding positive numbers has not been measured together in the give-N task, and theoretically, it is possible that they are independent (i.e., knowledge on positive numbers in itself cannot predict zero-knowledge). Importantly, in the present study, zero-knowledge is measured with the same task and with the same criteria as positive integer-knowledge; thus, this method could serve as an appropriate starting point to categorize the children in terms of whether they know what "zero" refers to. The give-N task used in this study is similar to the give-a-number task as described by Wynn (1990) and Bialystok and Codd (2000), which is the consensually accepted tool to measure preschoolers' cardinality principle- and number-knowledge. At the end of the trial, the experimenter explicitly asked the child whether they were done with the task. This is essential in tasks including zero-trials since the child does not have to give any items.

When measuring their understanding of zero with the task, two versions of zero were applied. For the "natural" version of zero, we utilized the form that is used in everyday language: "Do not give any balls to the bird." For the "mathematical" version, we used the form that reflects the mathematical viewpoint: "Give zero balls to the bird." In Hungarian, nouns after a number word are always singular, e.g., "give zero *ball*" or "give two *ball*," therefore, plural vs. singular form did not influence how the children solved the task (See the questions in all tasks in Hungarian and in their English translations in **Table A1**).

Overall, the following six numbers were tested in the given order: 2, 0 (natural), 5, 3, 0 (mathematical), 4. The present order of the numbers was used to prevent the children from relying on an order of increasing number words in the task. The whole number series was repeated twice, resulting in 12 trials.

It has been argued that children solving tasks with numbers 4 or larger are exhibiting an understanding of the cardinality principle (Wynn, 1990, 1992; Condry and Spelke, 2008; Sarnecka and Carey, 2008). In this study, a child was categorized as knowing a number if both trials of that number were solved correctly and if known numbers were not given as a response to higher unknown numbers[2]. Following these results, children were categorized as CP-knowers if they could give the numbers 4 and 5 in both trials; otherwise, they were categorized as subset-knowers. Categorization based on this task was performed right after the child completed the task; this was done because the comparison, addition, and subtraction task stimuli (see below) depended on the children's number knowledge.

Independently of the previous categorization, children were categorized as "nothing-givers" if they correctly did not give anything in the natural-zero task in both trials and as "zero-givers" if they correctly did not give anything in the mathematical-zero tasks in both trials. (Note that, because the performance of the give-N task with positive numbers is well-known in the literature, children who successfully give a specific value are termed "knowers," such as one-knowers, subset-knowers, or CP-knowers. However, such knowledge in the literature is not available for zero; therefore, to highlight the fact that it is not clear whether children solving this task with zero really understand some key features of zero, we term such children as nothing-"givers" or zero-"givers," instead of using the term "knowers," i.e., we are referring to the performance in the task instead of the supposed knowledge of the child).

Comparison

The aim of this task was to test whether the children knew the position of zero among other numbers, therefore, whether they are able to handle zero as efficiently as positive numbers (Aim 1) and to investigate whether the form of the task (natural verbal vs. mathematical verbal vs. non-verbal) influences their performance (Aim 2). Additionally, the number-knowledge groups, as identified by the give-N task, are compared in the task (Aim 3), and comparison operation performance is contrasted with the meta-knowledge tasks (Aim 4). In the task, the children either saw two sets of objects or heard two numbers, and had to choose the larger one.

To test whether the children understood the verbal description of the task, we used both verbal and a non-verbal object versions. In the object version, two sets of balls were placed on opposite sides of a table, and the question was, "On which side can you see more?" The number of the objects in a set was not named by the experimenter. In the case of the zero value, the appropriate side of the table remained empty. In the verbal version, no objects were used, and the question was, "Which one is more, the x or the y?" (The Hungarian translation of "which one" does not include the word "one," thus, this part of the question would not have confused the children.) In the verbal condition, the number zero was labeled as either "zero" (mathematical version) or "nothing" (natural version).

If the children understood the cardinality principle as measured by the give-N task, the following 12 number pairs were tested in the given order: 3–2, 4–1, 2–4, 3–zero, 2-nothing, 1–5, zero–4, 1–3, 2-zero, nothing-4, 2–1, and zero–1 (4 pairs with

[2]Number knowledge as measured with the give-N task can be calculated with several methods. Here, we consider two alternative evaluation methods, and report that these methods gave the same results as the method we reported in the main text. (1) In most published works, a number is judged to be known if the correct response rate is not lower than 66%; here, we used the 100% threshold value. Because we asked all numbers twice, the correct response could be only 100, 50, or 0%. The 100% criterion would underestimate the child's number knowledge compared to the usual 66% criterion, while the 50% criterion would overestimate it. Because the correct response rate changes rapidly around the limit of the child's knowledge, the 100 and 50% percent criteria give similar results. Given our analysis and exclusions, in most of our analysis the 50% criterion categorizes only two children as CP-knowers who were subset-knowers with the 100% criterion. We reran all analyses with both criteria, and they gave the same pattern of significant and non-significant results. For the sake of simplicity, we only present the results with the 100% criterion analysis here. (2) We also used an alternative Bayesian calculation method to specify whether someone is a subset-knower or a CP-knower (Negen et al., 2011), although, to our knowledge, the validity of this method has not been tested so far. It gave us a categorization result that was between the results of the previous 100 and 50% methods, thus, the results of this categorization are not presented here either.

"zero," 2 pairs with "nothing," and 6 pairs with only positive values). Otherwise, the following 10 number pairs were tested in the given order: 3–2, 2–1, 1–2, 1–zero, 2–nothing, 1–3, zero–2, 1–2, 2–zero, and nothing-1 (3 pairs with "zero," 2 pairs with "nothing," and 5 pairs with only positive values). The use of the two series for the two groups ensures that (a) the children see tasks only with a number range corresponding to their capability, while (b) it is possible to measure a relatively wide number range in the CP-knower group. Note that the two series do not prevent us from investigating the main questions: whether zero is processed as correctly as positive numbers and whether linguistic forms have an effect on performance, because no direct comparison of the two groups in terms of positive number performance is required, and only tasks or conditions within the groups are compared. Because, in the object condition, the "zero" and "nothing" versions mean the same stimulus (i.e., missing objects), only trials with "zero" were tested.

In the analyses, the percentage of correct responses in the "zero," "nothing," and positive trials was used as the task's performance index.

Addition and Subtraction

As in the comparison task, the aim for the addition and subtraction tasks was to test whether children can handle zero as efficiently as positive numbers in arithmetic operations (Aim 1), whether the form of the task (natural verbal vs. mathematical verbal vs. non-verbal) influences their performance (Aim 2), whether the number-knowledge groups as identified by the give-N task perform differently in the arithmetic tasks (Aim 3), and whether zero handling in arithmetic task performance is better than meta-knowledge task performance (Aim 4). In the arithmetic tasks, the children had to add or subtract two numbers and say the result.

To test whether the children understood the verbal description of the task, the tasks were either in verbal form or shown with objects. In the verbal form, either the "natural" or "mathematical" form of zero was used. In the object version, while the task was also explained in verbal form (in the case of zero, both in the "natural" and "mathematical" forms), the operands of the task were shown with arrays of balls: one operand on one side and the other operand on the other side of the table. In the case of zero, the appropriate side of the table remained empty.

If the child understood the cardinality principle as measured by the give-N task, the following 15 operations were tested in the given order: 1 + 1, 3 + 1, 2 + 0, 1 + 2, 0 + 3, 2 + nothing, nothing + 3; 2–1, 4–2, 3–0, 3–1, 2–0, 2–2, 3–nothing, and 2-nothing (4 tasks with "nothing," 4 tasks with 0, and 7 tasks with only positive values). Otherwise, the following 13 operations were tested in the given order:1 + 1, 2 + 1, 2 + 0, 1 + 2, 0 + 1, 2 + nothing, nothing + 1; 2–1, 2–0, 1–0, 1–1, 2–nothing, and 1-nothing (4 tasks with "nothing," 4 tasks with 0, and 5 tasks with only positive values only). The motivation for using the two series was the same as in the comparison task.

The task was embedded in a small story. In the addition task, the following story was told to the children: "The bird had x balls. The dog had y balls. How many balls do they have altogether when they play together?" In the subtraction task, the story was as follows: "The bird had x balls. Then, the bird gave y balls to the dog as a present. How many balls is the bird left with?" In the "natural" linguistic version, we applied forms that are commonly used in everyday speech, e.g., "The dog didn't have any balls." (In Hungarian, it is not possible to use a version that is close to the English "give no balls," because, in Hungarian, the predicate is negated, and the results are similar to "Do not give none balls to the dog.") In the "mathematical" linguistic version we used a form that reflects the mathematical viewpoint, e.g., "The dog had zero balls."

One may ask whether the "natural" versions of these tasks really measure numerical abilities or, instead, measure some other abilities. For example, it is possible that the children used a non-numerical concept of nothing to solve the tasks instead of the numerical concept of zero. However, it is important to understand that, for a preschooler, both the concept of nothing and the concept of zero are appropriate to solve numerical tasks with zero. At their age, preschoolers can solve only a few numerical tasks: comparison, addition, and subtraction (Levine et al., 1992); in all of these tasks, both the concept of nothing and the concept of zero give the same correct result. Therefore, a correct or erroneous numerical task cannot differentiate in a simple way whether the concept of nothing or the concept of zero was used; consequently, the question of whether any of these concepts promote different strategies is not testable with the current methods.

In the analyses, the percentage of correct responses in the "zero," "nothing," and positive trials was used as the task's performance index.

Smallest Number

The children were asked what the smallest number is. The aim of this task was to find out whether children regard zero as the smallest number and whether their performance on this task strongly correlates with the operations tasks (Aim 4). The task is similar to the task utilized in Wellman and Miller (1986). In the analysis, the given number was used directly. The task was applied as the first task of the session so that the children's responses could not be influenced by other tasks, which might have taught them about the number zero (See further details in the order of the tasks part below). Furthermore, because several of the other tasks could affect this knowledge, we repeated this task at the end of the session to determine whether the children' zero-knowledge had changed as a result of new knowledge they had acquired during the testing process.

Is It a Number?

The children had to categorize whether numbers as well as other things are numbers. The children were verbally asked, "Is the . . . a number?" The aim of this task was to study explicitly whether children regard zero as a number and whether their performance on this task strongly correlates with the operations tasks (Aim 4). To determine whether the children understood this categorization task, additional numerical and non-numerical words were used to validate the task. The following six words were tested in the given order: three, two, nothing, kitten, pop (sound), and zero. In the analysis, the given responses for the six trials were used directly. Similar to the smallest-number task, this task was presented at the beginning of the session, and it was repeated at the end of the session.

Order of the Tasks

Because some of the tasks could provide information about the meaning and use of the number zero to the children, we set the order of the tasks based on their potential to modify the responses of the children in later tasks. Therefore, the tasks concerning the status of zero were tested first. Additionally, since tasks including both verbal and object versions at the same time could teach the children the meaning of zero, we placed them at the end of the session. Finally, because the presented stimuli in the comparison, addition, and subtraction tasks depended on the number knowledge of the children, the give-N task preceded the comparison, addition, and subtraction tasks. Thus, the following order for the tasks was applied: smallest number, is-it-a-number, give-N task, comparison, verbal addition and subtraction, object addition and subtraction, smallest number (repeated), and is-it-a-number (repeated).

Procedure

After receiving the written consent of the parents and verbal consent from the children themselves, the measurement took place in a separate room in a building at their preschool. Data collection for a single child took approximately 30 min, and required a single session (See the order of the tasks above at the end of the tasks section).

Analysis Plan

To introduce the main contrasts and tests as they pertain to the tasks and aims, here, we provide a summary of the main analyses applied in the results section. Based on the give-N task, 2 × 2 categories are formed based on whether a child is a "zero"-giver or "zero"-non-giver and whether a child is a subset-knower or a CP-knower. In the comparison, addition, and subtraction tasks, (a) to see whether zero is more difficult to handle than positive numbers, correct response percentages are compared between tasks including zero and tasks including only positive numbers, (b) to investigate the effect of language, correct response percentages are contrasted in the verbal vs. non-verbal conditions and in the natural linguistic version vs. the mathematical linguistic version, and (c) to investigate the effect of number knowledge measured with the give-N task, correct response percentages are compared between subset-knowers and CP-knowers. In the smallest-number task, the distribution of the given responses are analyzed, and the frequencies of responses are compared between the different groups of number knowledge (i.e., subset-knowers vs. CP-knowers, and zero-givers vs. zero-non-givers). In the is-it-a-number task, correct response percentages are compared between non-numbers, positive numbers, and zero conditions, and this performance is compared between the groups of number knowledge (See more details on the specific analyses in the relevant results sub-sections).

RESULTS AND DISCUSSION

To make the results, their interpretation, and their connection to the specific aims easier to follow, this section groups the results by tasks and not by aims. Still, for all tasks, specific sets of their analyses are labeled by specific aims. In most task sections, the main results are described first; this is followed by relevant inferential statistics; and the section closes with the discussion of the results.

Groups Based on Number Knowledge

Giving Positive Numbers

First, with the give-N task, we specified the number knowledge level of the children (subset-knowers vs. CP-knowers). This categorization was used in the following tasks to contrast the children based on their number knowledge. Twenty children understood the cardinality principle (i.e., they could solve the give-N tasks perfectly for the numbers between 2 and 5), while 20 children had not reached yet this phase, i.e., they were subset-knowers. Two children from the latter group could not solve the give-2-balls task, and because the number one was not measured in the task, they could either be one-knowers or pre-knowers. As we could not specify whether these two children were subset-knowers or pre-knowers, we excluded them from further analysis. Among the remaining 18 subset-knowers there were seven two-knowers, ten three-knowers and one four-knower.

Giving "Nothing" and "Zero"

Second, independent of the previous number-knowledge categorization, we specified whether the children could solve the tasks involving the natural version of zero (nothing-givers) and the mathematical version of zero (zero-givers). Practically all children (96%) understood the natural version of the give zero task ("do not give any balls"), while the mathematical version ("give zero balls") proved to be more difficult (only 45% of the children could give "zero"). None of the nothing-givers or the zero-givers solved the task by adding zero accidentally for unknown numbers because zero was never given when a positive number was asked by the experimenter.

This difficulty with the mathematical version of zero could be rooted either in not knowing the word "zero," or in the unusual and unnatural form of the task (i.e., giving something which is nothing). As for the former explanation, notes by the children show that at least some of them did not know what the word "zero" refers to; examples include, "What does zero mean?," "I cannot count up to zero," "That would be too much for me," and "Zero is hundred."

Relation of Giving Positive Numbers and Zero

Computationally, giving positive numbers and giving "nothing" or "zero" could be independent. The following analysis investigates whether empirical data support independence. The present data show that most children could solve the natural-zero task, independent of whether they were subset-knowers or CP-knowers (92% of subset-knowers and 100% of CP-knowers), meaning that giving "nothing" is independent of whether a child is a subset-knower or CP-knower. Additionally, while giving "zero" depends on their positive number knowledge (63% of CP-knowers and 25% of subset-knowers gave zero successfully) giving zero is not strictly connected to such knowledge: Not all CP-knowers could give zero, while some of the subset-knowers could. Note that this result is unlikely to be random noise. For example, in our sample, the CP-knowers could give all positive numbers (i.e., between 2 and 5) successfully on two occasions, but they were unable to do so successfully with "zero." It is also noteworthy that even some of the subset-knowers understood the word "zero." All of the subset-knowers solving the mathematical version of the give-zero task were "three"-knowers.

The previously described effects were analyzed with a 2 (subset-knower group vs. CP-knower group as a between-subjects factor) × 2 ("natural" vs. "mathematical" versions of the zero task as a within-subjects factor) ANOVA on the proportion of correct responses: A main effect of number knowledge [$F(1, 36) = 7.59$, $p = 0.009$, $\eta_p^2 = 0.174$, better performance in the CP-knower group], and a main effect of linguistic version [$F(1, 36) = 46.62$, $p < 0.001$, $\eta_p^2 = 0.564$, better performance with the "nothing" version] were found to be significant; a tendency in the interaction [$F(1, 36) = 3.66$, $p = 0.064$, $\eta_p^2 = 0.092$, CP-knowers were more successful in giving "zero" than the subset-knowers] was also found.

Groupings for the Following Analyses

Positive number knowledge (whether a child is a subset-knower or a CP-knower) is used throughout the following tasks to investigate whether this knowledge influences understanding zero (Aim 3). Because giving "zero" is independent of their positive number knowledge and because giving "zero" may reflect their ability to understand the label "zero," the children were also categorized by this ability. A child was categorized as zero-giver, if the mathematical version ("zero") of the task was solved correctly in both trials. We could not rely on the natural version ("nothing") of the task because this task seemed to be trivial for preschoolers, resulting in a ceiling effect.

These two orthogonal dimensions created four groups of children that were used in the subsequent analyses (**Figure 1**): 4 subset-knower and zero-giver (4.08 years mean age 0.26 SD), 14 subset-knower and zero-non-giver (3.8 years mean age, 0.4 SD), 12 CP-knower and zero-giver (4.33 years mean age, 0.55 SD), and 8 CP-knower and zero-non-giver (4.06 year mean age, 0.37 SD). A child is categorized as a zero-giver if they can correctly solve the mathematical versions of the give-zero task.

Note that, in the operations tasks, because of some missing data the subset-knower—zero-giver group is excluded, therefore, only 3 groups are compared. However, for the meta-knowledge tasks all 4 groups are analyzed. Also note that, because there may be an interaction effect in their positive number knowledge and giving zero, in the upcoming analyses, the 4 (or 3) groups will be handled as 4 (or 3) independent groups and not as 2 × 2 factors.

Operations With Zero: Comparison, Addition, and Subtraction

These tasks investigated (1) whether the value of zero is more difficult to handle than positive integers, (2) the effect of language (a) contrasting verbal and non-verbal versions, which could reveal whether the children have either linguistic or conceptual problems with the zero value, and (b) contrasting the natural linguistic version (e.g., "add nothing to three"), and the mathematical linguistic version (e.g., "add zero to three"), and (3) the differences between CP-knowers and subset-knowers.

Because of a data collection problem, the data of 2 out of 4 subset-knower and zero-giver children's data were not available in these tasks. Since the data of the 2 remaining children may be misleading due to the extremely small sample size, this whole group was excluded from the analysis.

The results are presented as population estimations of the mean correct response proportions (**Figure 2**) and hypothesis tests of those values (**Table 2**). Different analyses were run for the three tasks (comparison, addition and subtraction—see rows in **Table 2**) and for the object and verbal versions of the tasks (see columns in **Table 2**). Within these analyses, the number-knowledge groups (x axes on **Figure 2**; subset-knower and zero-non-giver, CP-knower and zero-giver, CP-knower and zero-non-giver) and number types (columns in **Figure 2**; positive values, zero, and nothing, except in the comparison object version, where distinguishing nothing and zero would not make sense) were used. Mixed ANOVAs applied number knowledge as a between-subject factor and number types as a within-subject factor.

Object Version

First, in the object versions of the tasks, overall, zero was not more difficult to handle than positive values (left in **Figure 2**; Aim 1): In the appropriate ANOVAs, the main effect of the number or operand type was not significant in the comparison and addition tasks, and the interaction was not significant in either of the tasks (left in **Table 2**).[3] The only positive exception is that subtracting "nothing" was easier than subtracting positive numbers (a significant main effect of operand type in the object subtraction task; see **Table 2**), which is reasonable, since, in a relatively difficult subtraction task, it is easier to do nothing than to subtract a positive value. Thus, these results show that handling zero is not harder than other numbers for preschoolers in the object version of the tasks. This means that preschoolers understood the value of zero conceptually. This shows that comparison, addition and subtraction operations with zero are available as soon as one understands positive integers.

Note that the verbal variations of "nothing" and "zero" did not cause any significant effects (no main effect of number or operand type in comparison and addition tasks and no significant *post hoc* difference between those values in the subtraction task; see **Table 2**), which is understandable in these tasks where the verbal form could be complementary to the object based presentation (Aim 2). An early ability to handle zero as efficiently as positive numbers in these operations is true for both CP-knowers and

[3] In the figures and in the appropriate ANOVAs the critical information is the main effect of number type (positive numbers vs. zero vs. nothing), reflecting an overall effect of zero, or the interaction between number type and number-knowledge groups, reflecting a group specific effect of zero. However, the main effect of groups is not relevant because (a) differences between the groups only show whether some groups perform better in general and not the relative difference between zero and positive integer-related performance, and (b) the CP-knower and subset-knower groups received different tasks (see section "Materials and Methods").

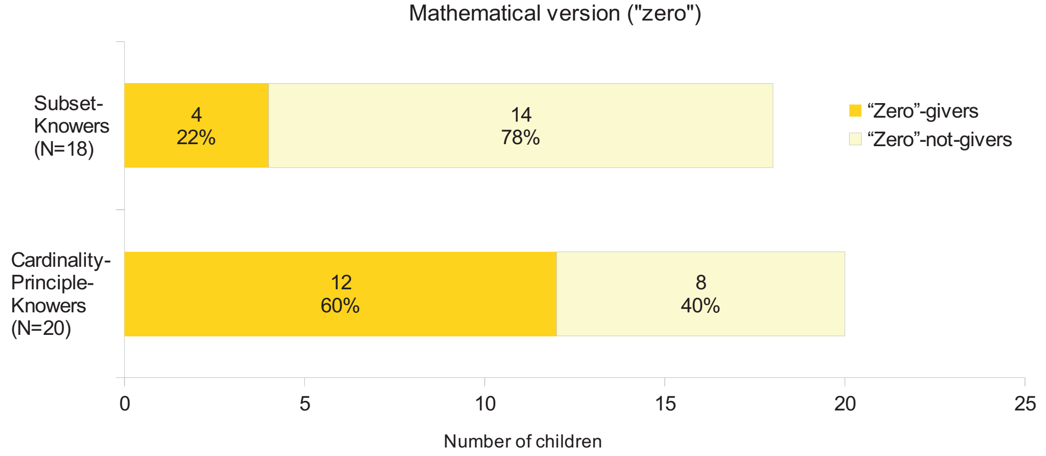


FIGURE 1 | Number of the children who understood the mathematical ("giving zero") version of the zero task.

subset-knowers as reflected in the same pattern in both groups (no interaction in any of the tasks; see **Table 2**; Aim 3).

Verbal Version

Second, in the purely verbal versions of the tasks, the zero value with the natural linguistic version ("nothing") was not more difficult to handle than the positive values (right in **Figure 2**): No main effect of number or operand type in comparison or addition tasks that is caused by the difference of "nothing" and positive number performance (right in **Table 2**; note that the significant difference of the mathematical version ("zero") is not a relevant contrast here). As in the case of the object version of the task, subtraction with "nothing" shows a positive exception, as it is easier than subtraction with positive numbers (significant main effect of operand type in the subtraction task; see **Table 2**). Again, these results show that understanding zero is available as soon as one understands positive integers (Aim 1).

However, handling zero values with the mathematical linguistic version ("zero") revealed difficulties (significant main effect of number or operand type in all tasks; see **Table 2**; Aim 2). Importantly, this difficulty was seen mainly in the zero-non-giver groups, but not in the zero-giver groups (significant interaction in the comparison task, see **Table 2**; and see similar patterns in the population estimations in all of the three tasks, see **Figure 2**). This pattern means that in a trivial way, children who do not know what the word "zero" refers to cannot solve the tasks including that unknown word. This interpretation is confirmed by the fact that these tasks can be solved when the task is also supported by object demonstrations.

Finally, the relative difficulty of handling zero and the linguistic form did not differ between subset-knowers and CP-knowers, as reflected in the similar relative patterns across the number-knowledge groups (see the similar patterns in parameter estimations between the subset-knowers and zero-non-givers and the CP-knowers and zero-non-givers, **Figure 2**; Aim 3).

To summarize, the verbal version of the operations tasks also confirms that preschoolers can handle zero as efficiently as positive numbers, with the trivial exception of children who do not know what the word "zero" refers to. Since the same tasks were solved efficiently in the object task, it confirms that the difficulty was linguistic in nature and not conceptual. Again, no difference between subset-knowers and CP-knowers could be found in these contrasts.

Alternative Interpretations

Some alternative interpretations of these data could be considered. One may raise that, when solving addition or subtraction tasks with zero, children cannot solve the task and, instead, simply repeat the last operand, which would sometimes result in the correct solution, e.g., in the task 0 + 3, repeating the last operand is the correct result. However, our results are not in line with this possibility. First, if the children did not know zero and only used this mechanically incorrect solution, they should have given incorrect responses in the comparison task as well, which was not the case. One may think that the children could have used the appropriate solution in comparison, and it is only addition and subtraction where they used the incorrect method. Nonetheless, this account cannot explain that if the children know how addition and subtraction work (as seen in the positive operand tasks) and if they know how zero can be ordered relative to other values (as seen in the correct zero-comparison task), why they do not use these pieces of knowledge in a zero-arithmetic task. Second, in these tasks, this last-operand repeating strategy would give the correct result only in half of the present additions, e.g., for 0 + 3, the correct result can be given, while for 2 + 0, an incorrect result would be given, and none of the subtraction tasks can be solved, e.g., for 2–0 an incorrect result could be given (see the specific stimuli in the Tasks section). However, in the addition task, the addition performance is higher than the 50% performance that could be expected by this alternative interpretation. Still, one may think, that in those tasks, children do not use the last operand, but they use the known operand. However, in other tasks (e.g., the comparison tasks) the same children handled zero appropriately, so it is less likely that they do not know zero.

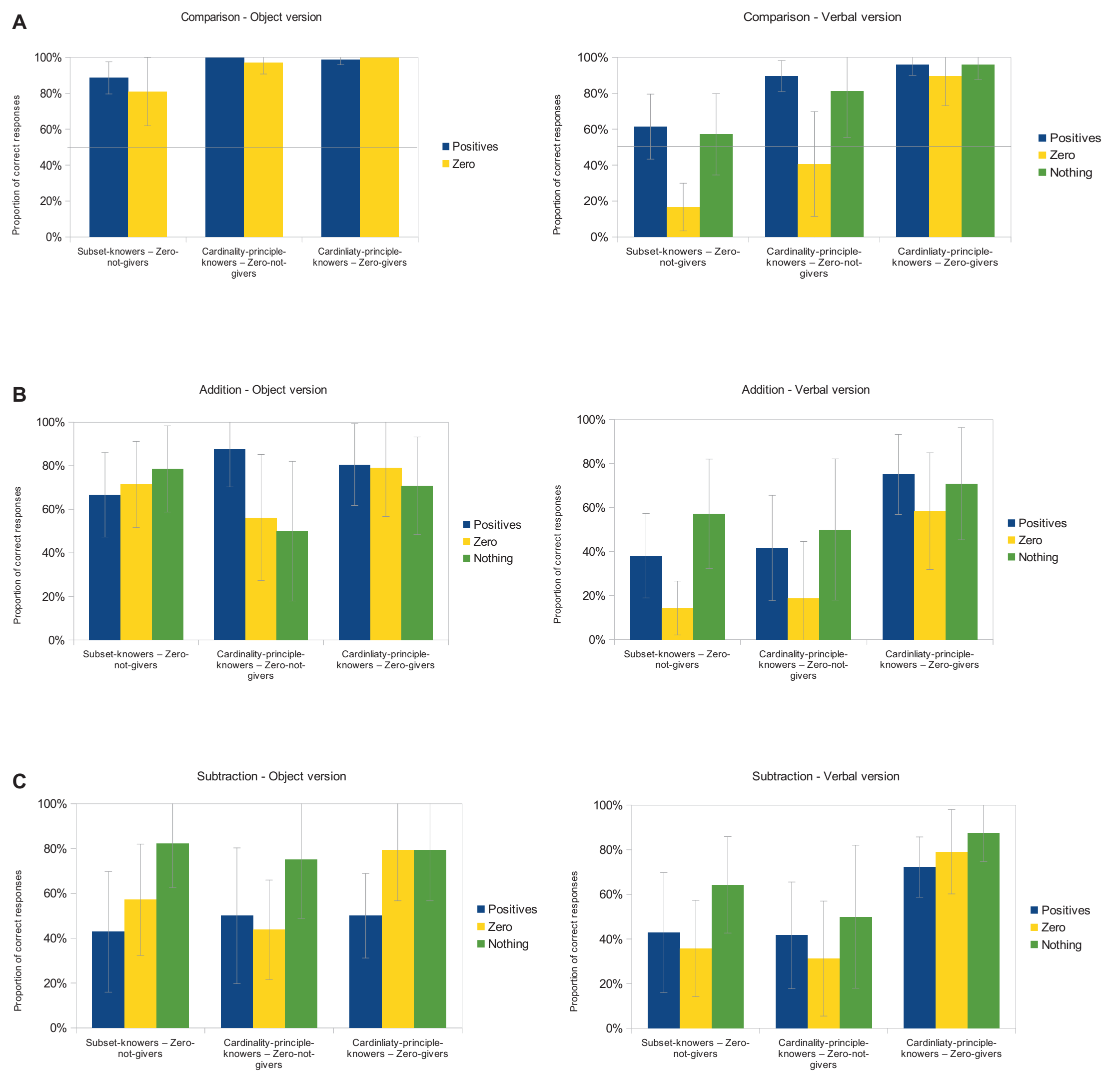


FIGURE 2 | Proportion of correct responses in the comparison **(A)**, addition **(B)**, and subtraction **(C)** tasks in the non-verbal (left) and verbal (right) versions with positive integers, zero, and nothing (columns) within the different number-knowledge groups (categories on x axes). Error bars show 95% confidence interval. Horizontal line at 50% in the comparison task shows the random choice level.

Another potential incorrect heuristics is that the children could compare the numbers instead of adding or subtracting them, which is a strategy that could result in correct responses in the tasks that include zero. However, this strategy would result in incorrect responses in the tasks that include only positive numbers, and such incorrect responses were not observed; thus, most preschoolers did not use this strategy either. Overall, if all tasks and results are considered together, one can conclude that the children handle zero on the same level as they handle positive numbers.

Meta-Knowledge of Zero

Smallest Number

Most children thought that the smallest number was "one" (**Table 3**). Although some zero-givers proposed "zero" as the smallest number, even most of the zero-givers thought that the smallest number was "one." None of the children proposed a negative number as an answer, although this happened once in one of our pilot studies.

These statements are supported by a chi-squared test on the proportion of responses (0; 1; other numbers; nothing; does not know), which showed a significant effect [$\chi^2(4, N = 38) = 26.47$, $p < 0.001$], reflecting a high proportion of "one" responses. A chi-squared test on the responses between the four number-knowledge groups revealed a significant difference [$\chi^2(12, N = 38) = 23.5$, $p = 0.024$], which most probably reflects the heterogeneous responses in the subset-knower and zero-non-giver group and the relatively uniform "one" responses in the other groups.

At the end of the session (i.e., after solving all other tasks), the smallest-number task was repeated, with the answers basically the same pattern as at the beginning of the session.

TABLE 2 | ANOVA results of the operations tasks.

	Object version	Verbal version
Comparison	**Number type main effect:** ns **Group main effect:** $F(2, 31) = 4.13$, $p = 0.026$, $\eta_p^2 = 0.211$; the subset-knowers show worse performance compared to the other two groups, $ps < 0.037$ **Interaction:** ns	**Number type main effect:** $F(2, 62) = 14.13$, $p < 0.001$, $\eta_p^2 = 0.313$; the performance for "zero" was worse than for "nothing" or positives, both $ps < 0.001$ **Group main effect:** $F(2,31) = 18.95$, $p < 0.001$, $\eta_p^2 = 0.55$; all groups differed from each other, all $ps < 0.017$ **Interaction:** $F(4, 62) = 2.45$, $p = 0.056$, $\eta_p^2 = 0.136$
Addition	**Operand type main effect:** ns **Group main effect:** ns **Interaction:** ns	**Operand type main effect:** $F(2, 62) = 5.91$, $p = 0.004$, $\eta_p^2 = 0.016$; tasks with "zero" were harder to solve than the two other task types, both $ps < 0.02$, LSD-test. **Group main effect:** $F(2, 31) = 5.14$, $p = 0.012$, $\eta_p^2 = 0.249$; CP-knower and zero-giver group shows better performance compared to the other two groups, both $ps < 0.018$. **Interaction:** ns
Subtraction	**Operand type main effect:** $F(2, 62) = 7.84$, $p = 0.001$, $\eta_p^2 = 0.202$; the tasks with "nothing" were easier to solve than the two other task types, both $ps < 0.003$ **Group main effect:** ns **Interaction:** ns	**Operand type main effect:** $F(2, 62) = 3.13$, $p = 0.05$, $\eta_p^2 = 0.092$; the tasks with "nothing" were easier to solve than the tasks with positives, or "zero," both $ps < 0.05$ **Group main effect:** $F(2, 31) = 5.92$, $p = 0.007$, $\eta_p^2 = 0.276$; the CP-knower and zero-giver group shows a better performance compared to the other two groups, both $ps < 0.007$ **Interaction:** ns

All analyses were run on the correct response ratios. The following designs were applied: in the object comparison version, 2 (number types: pairs with positive numbers vs. pairs with zero as a within-subjects factor) × 3 (number-knowledge groups as a between-subjects factor); in the verbal comparison version, 3 (number types: pairs with positive numbers vs. pairs with "zero" vs. pairs with "nothing" as a within-subjects factor) × 3 (number-knowledge groups as a between-subjects factor); in the addition and subtraction tasks, 3 (operand types: positive operands vs. zero vs. nothing as a within-subjects factor) × 3 (number-knowledge groups as a between-subjects factor). Post hoc tests are LSD tests. Note that, from the viewpoint of the current tests, the main effect of number or operand types and the interactions are relevant, but not the main effect of groups.

When evaluating the possible dissociation of zero knowledge in the former comparison task and in the present smallest-number task (Aim 4), we should consider zero-givers (because zero-non-givers typically could not name zero) and CP-knowers (because, in the comparison task, we had no subset-knowers within the zero-givers). The smallest-number task is in clear contradiction with the comparison task: While in the comparison task, zero-giver CP-knower children knew that zero was smaller than one (83% correct response performance in the verbal version and 100% in the object version), they believed that one was the smallest number (only 25% of them believed that zero is the smallest number). This result repeats the pattern that can be seen in the data of Wellman and Miller (1986). One could argue that the children may have misunderstood the task. However, it is hard to imagine that they misunderstood the smallest-number task, since CP-knower children can understand conceptual properties of numbers, such as that number words for large numbers with an unknown position in the counting list are numbers (Lipton and Spelke, 2006), and they can understand how set size change runs parallel with the counting list steps (Sarnecka and Carey, 2008). A possible resolution for this result is that preschoolers believed that zero was not a number. The subsequent task more explicitly tested whether or not children regard zero as a number.

TABLE 3 | Proportion of different replies to "what is the smallest number?" in the four groups of number knowledge at the beginning of the session.

	Subset-knowers		Cardinality-principle-knowers	
	Zero-non-giver (*N* = 14)	**Zero-giver (*N* = 4)**	**Zero-non-giver (*N* = 8)**	**Zero-giver (*N* = 12)**
Zero	1 (7%)	1 (25%)		3 (25%)
One	2 (14%)	3 (75%)	6 (75%)	9 (75%)
Two	2 (14%)		1 (13%)	
Three	1 (7%)			
Five	1 (7%)			
Nothing	2 (14%)			
Does not know	5 (36%)		1 (13%)	

Contrasting the subset-knowers and the CP-knowers in this task (Aim 3), on the one hand, neither of these groups considered zero to be the smallest number. On the other hand, the subset-knowers gave more heterogeneous responses, which is in line with the fact that they experienced difficulties in solving the verbal comparison task (**Figure 2**). Overall, while the subset-knowers show a qualitatively different response in the smallest-number task compared to the CP-knowers, this difference is most probably related to their capability in comparison and not to their zero-knowledge.

Is It a Number?

Most groups thought that "two" and "three" were numbers, while "pop" (sound) and "kitten" were not numbers, although the subset-knower—zero-non-giver group showed an approximately random performance (**Figure 3**). This means that, except for the latter group, the preschoolers understood the task.

Critically, while the word "nothing" was not evaluated as a number, the status of "zero" is uncertain. This result is in line with the finding that even some of the adults question whether zero is a number (Wheeler and Feghali, 1983) and could be in line with our interpretation of the results from the smallest-number task, as described above.

Supporting these statements, a 4 (number-knowledge groups as a between-subjects factor) × 6 (words to evaluate as a within-subjects factor) ANOVA on the proportion of correct responses (not on the proportion of "yes" responses as displayed in **Figure 3**) revealed a main effect of groups [$F(3, 34) = 3.008$, $p = 0.044$, $\eta_p^2 = 0.21$], with the subset-knower and zero-non-giver group showing poorer performance compared to the other three groups (all $ps < 0.037$), and an interaction between the two factors [$F(15, 170) = 2.12$, $p = 0.011$, $\eta_p^2 = 0.158$]. To find the source of this interaction, a similar ANOVA was run, excluding the subset-knower and zero-non-giver group. The 3

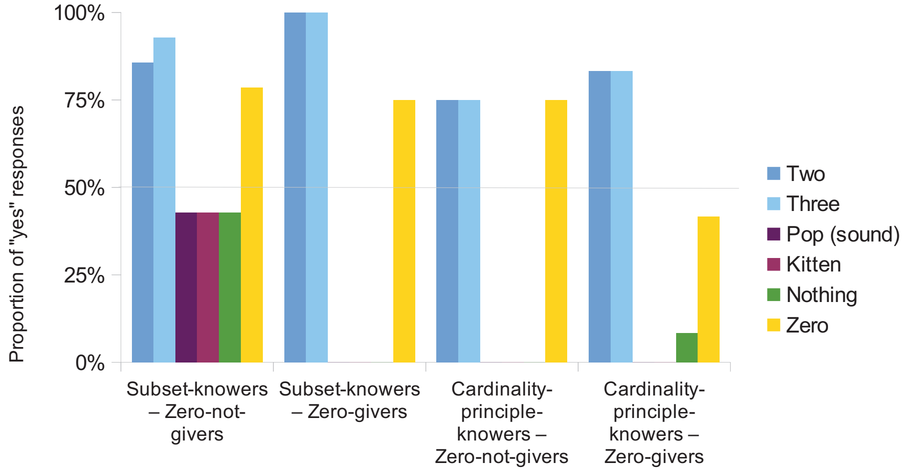


FIGURE 3 | Proportion of "yes" responses in the is-it-a-number task at the beginning of the session. Horizontal line at 50% shows the random choice level.

(groups) × 6 (words) ANOVA revealed only a main effect of words [$F(5, 105) = 3.34$, $p = 0.008$, $\eta_p^2 = 0.137$]. *Post hoc* LSD-tests showed that the error rate for "zero" was higher than for "pop," "kitten," and "nothing." Thus, in the three groups, evaluating "zero" was ambiguous, while the excluded subset-knower and zero-non-giver group showed random responses, reflecting that they did not understand the task.

Similar to the case of the smallest-number task, at the end of the session (after solving all other tasks), the is-it-a-number task was asked again, and the pattern of the responses did not change.

The results here show that, while the positive numbers are considered as numbers by the preschoolers, they do not think that the word "nothing" is a number, and they are ambiguous about the numerosity of the word "zero." (Note that here we ignore the results of the subset-knower—zero-non-giver group because, according to the data, this group did not understand the task.) The ambiguity of the word "zero" is partly understandable in the zero-non-giver—CP-knower group, since they can only guess. However, the result is more informative in the zero-giver groups since they could use the zero value correctly in numerical tasks even when the number was labeled as "zero." How can these differences be interpreted? There could be different factors that may influence children's decisions about numberness. First, all these words could be and were used by the preschoolers appropriately in the numerical tasks (at least in the CP-knower and zero-giver group), thus, the difference between the number status of the words cannot be caused by the children's numerical knowledge. Second, the word "nothing" can be used more generally in non-numerical cases and, additionally, unlike other number words, "nothing" cannot be used before nouns. These conceptual and linguistic differences may explain, why the word "nothing" is considered to be less numeric than the positive numbers and "zero." Importantly, none of these numerical, conceptual and linguistic considerations could explain why the word "zero" is considered to be less numeric than the positive numbers. Thus, we argue that the results reflect that preschoolers do not regard "zero" as a typical number. (Although the word "nothing" is also considered to be less numeric than the positive numbers, one cannot tell how strongly the mentioned conceptual and linguistic viewpoints influenced this decision and, consequently, how strongly the real perceived number status of "nothing" influenced the decision.) This could mean that either (a) at least some of the children do not regard "zero" as a number, or (b) judging the numberness of a value on a continuous scale (e.g., prototypicality of the value as a number) "zero" is not considered as a typical number, or (c) most children are confused about the number status of zero, and simply guessed in this task. To summarize, this result is consistent with the idea that children do not think that zero is a typical number.

Regarding the connection between operational knowledge and meta-knowledge (Aim 4), the same difference can be observed here as in the smallest-number task: While these children (and especially in the zero-giver–CP-knower group that may be our main interest) can handle zero in numerical operations, they are uncertain whether zero is a number. This finding is again in line with the additional phase we emphasized in the data from Wellman and Miller (1986): At a specific point in their development, preschoolers can handle numerical operations with zero, although they do not regard zero as a typical number.

When contrasting the subset-knower and CP-knower groups, the only difference is that the subset-knowers seem to be unsure about the meaning of the task, and seemingly subset-knower–zero-non-giver children gave responses around the 50% random level for all categories.

Reliability of the Data

Forty preschoolers participated in this study, who were categorized into four groups in the analyses, which resulted in rather small groups. Consequently, one may question how reliable our results can be. (1) Obviously, hypothesis tests control for small sample sizes: If the present results are significant, they are significant despite the small sample size. In other words, smaller samples need larger effect sizes to reach a significant result. This means that some of the current effects of interest are so large that even a relatively small sample is sufficient for a significant test or an appropriate power. (2) Moreover, a similar study was run before the present study. The present study is an improved version of our first measurement. The main differences between our previous and current measurements are that (a) the order of the tasks was modified, since, in the present version, all tasks that could potentially train the children were at the end of the session, (b) a few simple control tasks were removed, (c) there were minor improvements in the stimuli and instructions to have better control over the stimulus properties and the verbal conditions, and (d) the children were older, instead of 3- and 4-year-old children, 4- and 5-year old children participated. Importantly, the pattern of results was the same between the two studies: In the comparison and arithmetical tasks, zero was handled at the same level as positive numbers in both the subset-knower and CP-knower groups, even though the children were unsure about the number status of zero. Two notable differences that we found could readily be explained by the fact that the participants of the first measurement were older than the participants of the main measurement: The performance of the first measurement on the number status of zero was somewhat better than the performance of the sample in the present main study, and, in the pilot study, no zero-giver subset-knowers were found while there were only 6 subset-knowers out of the 36 participants. Overall, while single measurements could be unreliable (Open Science and Collaboration, 2015) and replications can be vital in obtaining reliable results, the present replication confirms that our findings are reliable. (3) Additionally, the smallest-number and is-it-a-number tasks were repeated within the session; as reported above, the results showed the same pattern, confirming again that the present results are reliable. (4) Moreover, the small size of the subset-knower, zero-giver group (4 children) cannot distort our results because their data in the operation tasks was not used at all, and almost all of our results rely on the data of the other three groups. (5) Finally, the results cannot be seen as a set of type-I and type-II errors or random noise because (a) similar and coherent patterns can be seen across independent tasks, (b) the results are coherent with many former reports, and (c) the results form a meaningful and coherent picture of the development of preschoolers' understanding of zero. With such a large number of tasks and statistical analyses, it is highly improbable that a small sample and related random variation could form a coherent picture as observed here. Overall, these considerations strongly suggest that the present results are reliable and reveal real developmental patterns.

GENERAL DISCUSSION

Our results shed light on several important aspects of preschoolers' understanding of zero. First of all, our data demonstrate that preschoolers understand the handling of empty sets in a numerical context and can appropriately apply it in various numerical tasks, such as giving a set with zero items, comparison, addition, and subtraction (Aim 1). Importantly, even subset-knowers can process empty sets appropriately: Their performance with empty sets was comparable with positive values if neither the mostly unfamiliar word "zero" nor the mathematical linguistic form was used (part of Aim 3). This result is in contrast with the results of Wellman and Miller (1986), who suggested that handling zero is difficult for preschoolers; this contradiction partly arises from the fact that, in that study, the children potentially could not understand the word "zero" and the mathematical linguistic form (see also the discussion of Aim 2 below). Additionally, while Wellman and Miller (1986) propose that children understand zero only when they know that zero is the smallest number, we argue that handling zero in numerical tasks is a sufficient criterion, while knowing that zero is the smallest number is knowledge that is irrelevant for evaluating preschoolers' operational capabilities (see also the discussion of Aim 4 below).

Second, while preschoolers can handle empty sets, they have difficulties with mathematical language (Aim 2). One component of this linguistic difficulty is the knowledge about the meaning of the word "zero." The children who did not know the word "zero" could not solve the tasks in which the number zero was denoted purely by this label, while they could solve the same task with different wording, suggesting that the difference is whether they know that the "zero" label refers to a concept that they can otherwise use. Relatedly, some children explicitly noted that they do not know the meaning of the word "zero." As a second linguistic component, the children had difficulty with the "mathematical" formulation of the tasks (e.g., "give zero balls to the bird"). Importantly, they could solve the task if (a) zero was denoted non-verbally or (b) the natural linguistic form of the zero-related statements (e.g., "do not give any balls to the bird") was used. This means that again, the difficulties described here are linguistic, not conceptual, in nature. It is not surprising that the children had problems with the mathematical formulations since this form is mostly used for mathematical and formal purposes, and they had probably rarely heard it before. Our results also showed that these language-related difficulties could be observed independent of whether a child is a subset-knower or a CP-knower (part of Aim 3).

Third, the children had problems with the number status of zero (Aim 4). First, they were unsure whether zero is a number. Second, they exhibited a contradiction: They thought that 0 is smaller than 1, but they thought that 1 is the smallest number (the same contradiction was apparent in the study of Wellman and Miller, 1986). This contradiction can be seen as another reflection of being unsure whether zero is a number: While 0 is smaller than 1 (nothing is less than something), 1 is the smallest number because 0 is not a number. Again, this pattern was independent of whether a child is a subset-knower or a CP-knower, although the subset-knowers had some difficulty in understanding the meta-knowledge tasks.

Note that, while the present study did not measure detailed sociodemographic characteristics, former numerical training, or other related properties that may qualify the present results, when comparable, our results are in line with the results of previous studies. These correspondences suggest that the present findings are at least robust.

Overall, these results identify several more or less independent components of the zero-knowledge. The first component is operational knowledge, i.e., whether children can use zero in tasks, such as giving a set of objects, comparing values, or adding and subtracting numbers. The second component is linguistic knowledge, including whether children know that the label "zero" is used to denote an empty set and whether they know the specific (and somewhat contradictory) mathematical formulation (e.g., "add zero balls to the rabbit"). The third component is meta-knowledge: Whether children know that zero is a number. The present results demonstrated that these three components do not necessarily appear at the same time. Our data also demonstrate that these pieces of knowledge do not depend strictly on whether a child is a subset-knower or a CP-knower.

Theoretical Account of the Development of Zero-Knowledge

What representations or accounts can explain this more detailed picture of the development of preschoolers' understanding of zero? First of all, the linguistic effect is not surprising in the sense that linguistic knowledge of zero has an effect on how zero is handled in verbal tasks. From a theoretical viewpoint, what is more interesting is the different development of operational and meta-knowledge of zero. How is that children who do not think that zero is a typical number still can solve numerical tasks with zero correctly? Several explanations offer a solution to this problem.

A group of explanations may suggest that linguistic or cultural factors cause the difference between zero and positive integers in some tasks and, relatedly, they explain why operational and meta-knowledge of zero dissociate. However, we argue that these explanations cannot account for the present results. (1) An explanation posits that the special status of zero is due to its infrequent use in everyday language (Dehaene and Mehler, 1992). However, after understanding the cardinality principle children understand some common properties of numbers, even if that number is beyond the limit of their counting list (Lipton and Spelke, 2006). Critically, the frequencies of such numbers are comparable to the frequency of zero. Consequently, the relatively low frequency of the word "zero" cannot explain why children do not regard zero as a number, because, while a rare number (i.e., zero) is not regarded as number, other rare numbers (e.g., 20 and 40) are regarded as numbers. (2) Another example of a linguistic explanation posits that zero is categorized incorrectly because empty sets are handled linguistically in a different way than positive numbers: The word "nothing" is commonly used instead of "zero" and "natural" sentences are used instead of the "mathematical" versions. However, zero-givers (those who are familiar with these linguistic forms) are as unsure of the number status of zero as zero-non-givers; therefore, even if these rare linguistic forms are known, zero still can be thought as a non-number. Additionally, in a similar case of the number one, although children learn the distinction between one and many (i.e., other positive numbers) grammatically, which helps them to learn the meaning of "one" (Carey, 2009), they still think that one is a number. Thus, in the case of this linguistic difference, the relevant value (i.e., one) is not categorized differently than larger natural numbers. (3) A final possible linguistic-cultural explanation proposes that a counting list usually starts with one, which seemingly excludes zero from the set of numbers (Merritt and Brannon, 2013). However, at least in some cases, this issue is not rooted in the counting list. For example, some adults also think that zero is not a number, and their justifications are not related to the counting list: They say that zero is not a number because "a number is the abstract value of the quantity of a set" or because "it has no value" (Wheeler and Feghali, 1983). Additionally, it is unclear why this linguistic phenomenon would have an effect on categorization if the previous and potentially stronger linguistic effects (i.e., the first and second points of this paragraph) do not play a role in the categorization of zero. Importantly, none of these linguistic-cultural explanations can account for the fact that zero is handled differently only in some tasks, while it is handled similarly to positive integers in other tasks. To conclude, although we cannot entirely exclude all linguistic or cultural influences, these explanations cannot fully account for the dual nature of zero.

A second group of explanations supposes that there are representational causes why zero is not regarded as a typical number. (1) One can imagine that the representation supporting positive values cannot store the value of zero; hence, zero must be stored in a different system. The most frequently referred-to model suggests that semantic numerical processing is driven by the Approximate Number System (ANS) which represents numbers in an imprecise way (Moyer and Landauer, 1967; Feigenson et al., 2004; Merritt and Brannon, 2013). This system may include zero, although its ability to do so is debated (Dehaene et al., 1993; Dakin et al., 2011; Merritt and Brannon, 2013; Ramirez-Cardenas et al., 2016) (See more details on this system in **Supplementary Material**). Some part of the results could readily be explained by this model: In the operation tasks, even if the stimuli were presented symbolically, the imprecise representation may be handling the relatively small values used in the present experiment. However, it is not trivial to consider how the model accounts for the dual nature of zero. In the ANS framework, one may assume that while the positive numbers and, in some cases, zero value could be handled by this system (as in the comparison task), in some other cases, the zero value should be handled by another system (as in the number-status-of-zero task). However, it is not clear what this alternative system could be, and why the ANS sometimes could not handle zero. In other words, these suppositions are arbitrary, and the additional details were only created to account for these new results, while they are not supported by any other phenomena. Overall, it is hard to provide a coherent explanation of how the ANS could account for the special status of zero. (2) Another possible explanation suggests that understanding numbers relies on a conceptual understanding of the items (e.g., objects) in a set. For example, Carey (2004, 2009) proposes that children induce a conceptual understanding of how counting can specify the size of a set, which relies on set-templates in long-term memory; such templates are based on the activation of visual indexes. (It is important to note that it is not simply the visual index or Object Tracking System that supports the cardinality principle, because the visual index in itself is only a set of spatial indexes, which computationally is not capable of supporting conceptual understanding. Instead, some additional mechanisms, such as set-templates in the long-term memory, and other unspecified mechanisms are necessary to support a conceptual understanding of the cardinality principle.) Another model for explaining why sign-value notation numbers (e.g., Roman notation) are easier to understand than place-value numbers (e.g., Indo-Arabic notation) proposes that sign-value notations are more similar to a hypothesized item-based number representation in which the powers (e.g., ones, tens, hundreds, etc. in a 10-based system) are denoted by objects or groups (Krajcsi and Szabó, 2012). In both of these latter models, numbers could be considered as the properties of items or objects in a set; since there are no items or objects in an empty set, it would not make sense to talk about the property of the non-existing items, and, accordingly, zero could not be a number in this framework. The case of a property for a missing item is similar to the paradox introduced by Lewis Carroll: In his story, a smiling cat disappears, though the smile of the cat remains visible. In a similar way, the property (in this case, the numerosity) of the non-existent items is not meaningful, and the lack of items cannot be described in a similar way to a set of items. This model is also in line with the justifications provided by adults for why they do not regard zero as a number, i.e., they often state that numbers describe the items in a set (Wheeler and Feghali, 1983). Importantly, this view can explain why zero has a dual nature: Even though empty sets can be handled in numerical tasks that involve manipulating objects or items, the status of zero is special since it cannot represent the property of items that are not present, but it represents the lack of those items. Therefore, we conclude that a conceptual understanding of numbers as items or objects can explain why zero is handled in a special way.

Implications for Practice

The present study has some educational consequences. Teaching and understanding the properties of zero is difficult (Lichtenberg, 1972; Frobisher, 1999). Understanding some simple properties, such as the parity of zero and the number status of zero, can be problematic, even for mathematics teachers (Wheeler and Feghali, 1983; Hill et al., 2008).

Based on our results, it is possible to offer some educational recommendations for schools and even preschools. At the same time, the efficiency of these recommendations may also validate and test our description of the zero-knowledge development. Our recommendations are based on the fact that several independent components of zero-knowledge can be identified (namely, operational knowledge, linguistic knowledge and meta-knowledge) and that such components develop at different paces. (1) Based on the present results for operational knowledge, children could handle empty sets as soon as they can handle positive numbers in basic numerical operations; thus, no further effort is needed to introduce this concept to them. (2) However, they are likely unfamiliar with the mathematical linguistic formulations of the problem; thus, one can teach them these linguistic forms. (a) One linguistic component to teach is that the word "zero" means "nothing" or "doing nothing" in mathematical problem-solving. (b) Another linguistic component to teach is the mathematical form of sentences; thus, one can recite both the "natural" and "mathematical" versions of the task, stressing that the two versions refer to the same thing, e.g., "do not give any balls to Suzy Sheep; you can also say, give zero balls to Suzy Sheep." (3) Additionally, children do not understand the number status of zero, and they are unsure whether zero is a number. According to one possibility, one can explain to children that zero is also a number because we can use it in counting and other mathematical tasks. However, this solution raises a problem. If our explanation for the present results is correct, then children are not sure whether zero is a number because they consider that a number is a property of items, and zero describes the lack of items, which is a quantitatively different state. This "no items" state can handle empty sets in numerical tasks correctly, although it is not regarded as a number. Importantly, before school (and, in most educational systems, also in the first years of school), children have no experience that could demonstrate to them that the "no items" state is actually a number. For example, most adults know that zero is a number because it is simply another step on the number line between positive and negative integers or because, in written multi-digit calculations, zero is handled in a similar way as other digits. However, preschoolers do not have any experience that could reveal the numerical nature of the "no items" state because, for example, they do not know negative integers, or they cannot make multi-digit calculations. Thus, although one can explain that zero is a number because we use it in numerical tasks, this information still cannot be justified based on their experiences. Consequently, it seems more reasonable that it is unnecessary to provide any education on this issue before school, and it may be better to teach children about the number status of zero only when this information can be justified and can be built up in a reasonable way. To sum up, (1) preschoolers already know how to handle empty sets in basic numerical operations, although (2) mathematical language of referring to empty sets can be taught, while (3) teaching the number status of zero is probably unnecessary at this stage of their development.

CONCLUSION

In conclusion, preschoolers can handle zero on the same level as they can handle positive integers. Although the linguistic form can cause difficulties for them, this is independent of their conceptual understanding. However, preschoolers are unsure whether zero is a number. This may be caused by the set-based representation of numbers: Numbers can be the properties of items in a set, and, since an empty set does not include any items, zero cannot be a number in this view.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ELTE Eötvös Loránd University, Department of Cognitive Psychology. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors contributed to the design of the study, analysis of the data, and writing the manuscript.

FUNDING

The work was supported by the CELSA Research Fund (project no. CELSA/19/011) and by the National Research, Development and Innovation Fund of Hungary (project no. 132165) (AK).

ACKNOWLEDGMENTS

We would like to thank Rita Csikai, Lilla Hodossy, Andrea Kikilai, Orsolya Kis, Kinga Kreif, and Barbara Szűcs, for their help in collecting the data and their suggested modifications of the applied tasks. We also thank Eszter Szabó for discussing with us the development of zero-knowledge, Ágnes Lukács for her suggestions on terminology, and Réka Börcsök, Dávid Czikora, Lilla Hodossy, Nicole Janossy, Krisztián Kasos, Michal Pinhas, and Eszter Szabó for their comments on an earlier version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.583734/full#supplementary-material

REFERENCES

Baroody, A. J., Lai, M., Li, X., and Baroody, A. E. (2009). Preschoolers' understanding of subtraction-related principles. *Mathemat. Think. Learn.* 11, 41–60. doi: 10.1080/10986060802583956

Bialystok, E., and Codd, J. (2000). Representing quantity beyond whole numbers: some, none, and part. *Can. J. Exp. Psychol. Revue Canadienne Psychol. Exp.* 54, 117–128. doi: 10.1037/h0087334

Bulthé, J., De Smedt, B., Op, and de Beeck, H. P. (2015). Visual number beats abstract numerical magnitude: format-dependent representation of arabic digits and dot patterns in human parietal cortex. *J. Cogn. Neurosci.* 27, 1376–1387. doi: 10.1162/jocn_a_00787

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

Carey, S. (2009). *The Origin of Concepts (1st ed.)*. Oxford: Oxford University Press.

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

Condry, K. F., and Spelke, E. S. (2008). The development of language and abstract concepts: the case of natural number. *J. Exp. Psychol. General* 137, 22–38. doi: 10.1037/0096-3445.137.1.22

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., and Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proc. Natl. Acad. Sci. U S A.* 108, 19552–19557. doi: 10.1073/pnas.1113195108

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

Davidson, P. M. (1992). *Precursors of Non-Positive Integer Concepts*. Available online at: https://eric.ed.gov/?id=ED356146 (accessed June 19, 2017).

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and mental number magnitude. *J. Exp. Psychol. General* 122, 371–396. doi: 10.1037/0096-3445.122.3.371

Dehaene, S., and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition* 43, 1–29. doi: 10.1016/0010-0277(92)90030-L

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

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

Fischer, M. H., and Rottmann, J. (2005). Do negative numbers have a place on the mental number line? *Psychol. Sci.* 47, 22–32.

Frobisher, L. (1999). "Primary school children's knowledge of odd and even numbers," in *Pattern in the Teaching and Learning of Mathematics*, ed. A. Orton (London: Cassell), 31–48.

Gebuis, T., and Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *J. Exp. Psychol.: General* 141, 642–648. doi: 10.1037/a0026218

Grana, A., Lochy, A., Girelli, L., Seron, X., and Semenza, C. (2003). Transcoding zeros within complex numerals. *Neuropsychologia* 41, 1611–1618. doi: 10.1016/s0028-3932(03)00109-x

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study. *Cogn. Instruct.* 26, 430–511. doi: 10.1080/07370000802177235

Hughes, M. (1986). *Children and Number: Difficulties in Learning Mathematics*. Oxford: B. Blackwell.

Kaufman, J., Csibra, G., and Johnson, M. H. (2003). Representing occluded objects in the human infant brain. *Proc. Royal Soc. London. Series B: Biol. Sci.* 270(Suppl. 2), S140–S143. doi: 10.1098/rsbl.2003.0067

Krajcsi, A., Kojouharova, P., and Lengyel, G. (2020). Processing symbolic numbers: the example of distance and size effects. *Psyarxiv* [preprint]. doi: 10.31234/osf.io/5wzcx

Krajcsi, A., Lengyel, G., and Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Front. Psychol.* 7:1795. doi: 10.3389/fpsyg.2016.01795

Krajcsi, A., and Szabó, E. (2012). The role of number notation: sign-value notation number processing is easier than place-value. *Front. Psychol.* 3:463. doi: 10.3389/fpsyg.2012.00463

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

Le Corre, M., and Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395–438. doi: 10.1016/j.cognition.2006.10.005

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

Lichtenberg, B. P. (1972). Zero is an even number. *Arithmetic Teacher* 19, 535–538. doi: 10.5951/at.19.7.0535

Lipton, J. S., and Spelke, E. S. (2006). Preschool children master the logic of number word meanings. *Cognition* 98, B57–B66. doi: 10.1016/j.cognition.2004.09.013

Merritt, D. J., and Brannon, E. M. (2013). Nothing to it: precursors to a zero concept in preschoolers. *Behav. Proc.* 93, 91–97. doi: 10.1016/j.beproc.2012.11.001

Merritt, D. J., Rugani, R., and Brannon, E. M. (2009). Empty sets as part of the numerical continuum: conceptual precursors to the zero concept in rhesus monkeys. *J. Exp. Psychol. General* 138, 258–269. doi: 10.1037/a0015231

Moeller, K., Zuber, J., Olsen, N., Nuerk, H.-C., and Willmes, K. (2015). Intransparent German number words complicate transcoding – a translingual comparison with Japanese. *Front. Psychol.* 6:740. doi: 10.3389/fpsyg.2015.00740

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

Negen, J., Sarnecka, B. W., and Lee, M. D. (2011). An excel sheet for inferring children's number-knower levels from give-N data. *Behav. Res. Methods* 44, 57–66. doi: 10.3758/s13428-011-0134-4

Noël, M.-P., and Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number representation model. *Front. Hum. Neurosci.* 5:165. doi: 10.3389/fnhum.2011.00165

Open Science, and Collaboration. (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716

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

Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. *Curr. Biol.* 26, 1285–1294. doi: 10.1016/j.cub.2016.03.052

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

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

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

Varma, S., and Schwartz, D. L. (2011). The mental representation of integers: an abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition* 121, 363–385. doi: 10.1016/j.cognition.2011.08.005

Wellman, H. M., and Miller, K. F. (1986). Thinking about nothing: development of concepts of zero. *Br. J. Dev. Psychol.* 4, 31–42. doi: 10.1111/j.2044-835X.1986.tb00995.x

Wheeler, M. M., and Feghali, I. (1983). Much ado about nothing: preservice elementary school teachers'. Concept of Zero. *J. Res. Mathemat. Educ.* 14:147. doi: 10.2307/748378

Wynn, K. (1990). Children's understanding of counting. *Cognition* 36, 155–193. doi: 10.1016/0010-0277(90)90003-3

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

Wynn, K., and Chiang, W.-C. (1998). Limits to infants' knowledge of objects: the case of magical appearance. *Psychol. Sci.* 9, 448–455. doi: 10.1111/1467-9280.00084

Zuber, J., Pixner, S., Moeller, K., and Nuerk, H.-C. (2009). On the language specificity of basic number processing: transcoding in a language with inversion and its relation to working memory capacity. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

APPENDIX

TABLE A1 | The questions in all tasks, in Hungarian and in their English translations.

Hungarian version	English translation	English translation reflecting the Hungarian structure more strictly
Give-N task		
Non-zero version: Adj a madárnak . . . golyót.	Give . . . balls to the bird.	Give to the bird . . . ball (*singular*).
Natural zero version: A madárnak semennyi golyót se adj.	Do not give any balls to the bird.	To the bird none ball (*singular*) do not give.
Mathematical zero version: Adj a madárnak nulla golyót.	Give zero balls to the bird.	Give to the bird zero ball (*singular*).
Comparison		
Object version: Melyik oldalon van több?	On which side is there more?	On which side is there more?
Verbal version: Melyik a több, a . . . vagy a . . .?	Which one is more, the . . . or the . . .?	Which is the more, the . . . or the . . .?
Addition and subtraction		
Addition: Mennyi golyójuk van összesen, ha együtt játszanak?	How many balls do they have altogether when they play together?	How many ball (*singular*) they have altogether, if together they play?
Subtraction: Mennyi golyó marad a madárnak?	How many balls was the bird left with?	How many ball (*singular*) has left for the bird?
Smallest number		
Melyik a legkisebb szám, mi a legkevesebb?	What is the smallest number, what is the fewest one?	Which is the smallest number, what is the fewest?
Is it a number?		
A . . . az szám?	Is the . . . a number?	The . . ., is it a number?

frontiers in Psychology

ORIGINAL RESEARCH
published: 06 September 2021
doi: 10.3389/fpsyg.2021.697881

Neuromuscular Diseases Affect Number Representation and Processing: An Exploratory Study

Hendrikje Schmidt[1,2*†], *Arianna Felisatti*[1†], *Michael von Aster*[2], *Jürgen Wilbert*[3], *Arpad von Moers*[4‡] and *Martin H. Fischer*[1‡]

[1] Potsdam Embodied Cognition Group, Department of Psychology, University of Potsdam, Potsdam, Germany, [2] Center for Special Educational and Psychological Needs, German Red Cross Hospitals Berlin, Berlin, Germany, [3] Department of Inclusive Education, University of Potsdam, Potsdam, Germany, [4] Department of Pediatrics, German Red Cross Hospitals Berlin, Berlin, Germany

OPEN ACCESS

Edited by:
Ricardo Moura,
University of Brasilia, Brazil

Reviewed by:
Janet Hoskin,
University of East London,
United Kingdom
Léo Dutriaux,
University of Glasgow,
United Kingdom

***Correspondence:**
Hendrikje Schmidt
h.schmidt@drk-kliniken-berlin.de

† These authors share first authorship

‡ These authors share last authorship

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

Received: 21 April 2021
Accepted: 04 August 2021
Published: 06 September 2021

Citation:
Schmidt H, Felisatti A, von Aster M, Wilbert J, von Moers A and Fischer MH (2021) Neuromuscular Diseases Affect Number Representation and Processing: An Exploratory Study. Front. Psychol. 12:697881. doi: 10.3389/fpsyg.2021.697881

Spinal muscular atrophy (SMA) and Duchenne muscular dystrophy (DMD) both are rare genetic neuromuscular diseases with progressive loss of motor ability. The neuromotor developmental course of those diseases is well documented. In contrast, there is only little evidence about characteristics of general and specific cognitive development. In both conditions the final motor outcome is characterized by an inability to move autonomously: children with SMA never accomplish independent motoric exploration of their environment, while children with DMD do but later lose this ability again. These profound differences in developmental pathways might affect cognitive development of SMA vs. DMD children, as cognition is shaped by individual motor experiences. DMD patients show impaired executive functions, working memory, and verbal IQ, whereas only motor ability seems to be impaired in SMA. Advanced cognitive capacity in SMA may serve as a compensatory mechanism for achieving in education, career progression, and social satisfaction. This study aimed to relate differences in basic numerical concepts and arithmetic achievement in SMA and DMD patients to differences in their motor development and resulting sensorimotor and environmental experiences. Horizontal and vertical spatial-numerical associations were explored in SMA/DMD children ranging between 6 and 12 years through the random number generation task. Furthermore, arithmetic skills as well as general cognitive ability were assessed. Groups differed in spatial number processing as well as in arithmetic and domain-general cognitive functions. Children with SMA showed no horizontal and even reversed vertical spatial-numerical associations. Children with DMD on the other hand revealed patterns in spatial numerical associations comparable to healthy developing children. From the embodied Cognition perspective, early sensorimotor experience does play a role in development of mental number representations. However, it remains open whether and how this becomes relevant for the acquisition of higher order cognitive and arithmetic skills.

Keywords: spatial-numerical associations, numerical processing, mathematics, child development, embodied cognition, neuromuscular disease, spinal muscular atrophy, Duchenne muscular dystrophy

INTRODUCTION

Children born with neuromuscular degeneration are rare. Depending on the exact disease type, incidence rates between 1 and 5 out of 10,000 children have been reported in prevalence studies (for a review, see Deenen et al., 2015). Affected children tend to be looked at only deficit-oriented on their motor capacity. Both individual cognitive/psychosocial deficits and strengths might be neglected or not even detected in the first place. In particular, children and adolescents affected by spinal muscular atrophy (SMA) or Duchenne muscular dystrophy (DMD) suffer from loss of muscle strength and progressive motor impairments, with implications for daily physical, cognitive and social functioning (e.g., Billard et al., 1992). So far, cognitive consequences of inherited motor deficits have been explored through intelligence scales, such as Wechsler Intelligence Scales, Kaufman Assessment Battery, Standard and Progressive Matrices. These instruments cover verbal and non-verbal cognitive abilities and have revealed that the impact of the motor impairment on cognitive parameters depends on the genetic background and the specific neuromuscular disease, as described below.

In the present study, we will focus on one of these higher-level cognitive abilities, namely numerical cognition. Numeracy, the ability to comprehend and mentally manipulate quantitative non-symbolic and symbolic relations, is a key predictor for a healthy and successful life (Parsons and Bynner, 1997; Woloshin et al., 2001; Butterworth, 2018). In light of differences in arithmetic performances with respect to the two disease groups (Billard et al., 1992; von Gontard et al., 2002), we aim to answer the question of *how* numerical concepts are represented and processed in children with SMA and DMD. We begin with a brief summary of typical and atypical development of numerical cognition, followed by a description of these specific rare neuromuscular diseases, drawing attention to differences between the two medical conditions. Then we motivate the present study of numerical abilities in SMA and DMD from an embodied cognition perspective.

Typical and Atypical Development of Numerical Cognition

Number processing, calculation and math reasoning constitute a complex cognitive domain that developmentally relies on certain core abilities to discriminate concrete non-symbolic magnitudes, that are available quite early in life. The achievement of further domain-specific cognitive representations and abilities in the numerical domain, like symbolic number processing (linguistic, Arabic), the spatially oriented mental number line, arithmetic procedures and mathematical reasoning abilities, is dependent on and accompanied by the development of domain-general abilities, like sensorimotor and visuospatial integration, working memory, language, attentional and affective regulation. Therefore, reasons, symptoms and course of atypical development of numerical cognition are manifold and diverse (Bachot et al., 2005; von Aster and Shalev, 2007; von Aster et al., 2021). As DMD and SMA are diseases with similar motor but quite different cognitive and social-emotional outcomes, it will be of interest how children of both groups differ according to their embodied cognitive numerical representations and developmental patterns of domain-specific and domain-general intellectual functioning.

Spinal Muscular Atrophy

SMA is an autosomal recessive inherited disease. It is caused by mutations in the survival motor neuron1-gene (*SMN1*-gene) on chromosome 5. They result in less or total absence of the survival motor neuron-protein (SMN-protein) and only affect the anterior motor horn cell in the spinal cord. SMN-protein is essential for life, being necessary for information processing from the spinal cord to the muscles (Lefebvre et al., 1995). As a consequence, absence of the *SMN1*-gene results in general weakness with proximal muscles more affected than distal ones. Three types are distinguished in children and adolescents depending in the onset of the symptoms. Severity of disease depends on the amount of SMN-protein present and is negatively correlated with symptom onset (Calucho et al., 2018). If symptoms occur before adulthood, three types of SMA are distinguished: Children with SMA type 1 show symptoms within 6 months after birth; they will never be able to sit autonomously and their life expectancy without treatment is up to 2 years. In SMA type 2 symptoms occur after 6 months of age, allowing children to sit but never to stand or walk. Scoliosis is frequent and some patients need non-invasive ventilation. SMA type 3 is defined by sufficient muscle strength for walking without support and proximal weakness within the first two decades. People affected by SMA show normal sensory, cognitive, emotional and social functioning (Kolb and Kissel, 2015). Since 2017 and 2020 molecularly based medication has an essential impact on the course of the disease.

Duchenne Muscular Dystrophy

DMD is an X-chromosomal recessive disorder caused by mutations on the short arm of the X-chromosome and therefore is exclusively affected in males. It results in the reduction of dystrophin protein. Isoforms of dystrophin are expressed in various forms in different organs, e.g., skeletal muscles, cardiac muscle and the brain. Even if DMD patients achieve expected milestones of motor development, they experience progressive loss of muscular strength. First symptoms of muscle weakness are normally appearing from 18 months to 3 years. Muscle weakness is proximal announced and slowly progressive. Loss of gate without therapeutic support in 9–12 years of age followed by scoliosis, involvement of cardiac muscles. Affection of the respiratory system has a life-limiting effect (Spuler and von Moers, 2004). Between age 12 and 18, this progressive muscular weakening extends to the heart and respiratory system, to the point of provoking premature death (Schaaf and Zschocke, 2018).

Dystrophin is not only present in muscle cells but also plays an important role in brain development (see below).

Abbreviations: SMA, spinal muscular atrophy; DMD, Duchenne muscular dystrophy; SMAs, children with SMA; DMDs, children with DMD; SNAs, spatial numerical associations; RNG, random number generation; FODs, first order differences.

Indeed, deficient production of dystrophin isoforms has severe consequences in different areas of cognitive ability (Taylor et al., 2010), for example in short-term memory capacity (D'Angelo and Bresolin, 2006). Today there is no complete cure for DMD: Indeed, only motor aspects and not cognitive and social functioning can be improved by optimal medical care (e.g., through physiotherapy).

Important Similarities and Differences Between SMA and DMD

In the above sections we described the genetic conditions for motor impairment in SMA and DMD. Indeed, both children with SMA and DMD (hereafter SMAs and DMDs, respectively) suffer from progressive weakness in the proximal muscles, resulting in the inability to explore the environment without assistive equipment, e.g., wheelchairs or help from caregivers. Additionally, in both disease groups different medical treatments and therapies might impact cognitive functions.

Although DMD and SMA patients both have severe motor impairments, their cognitive and psychosocial functioning differs. Consider first DMDs, where the lack of dystrophin in the hippocampus and the cerebellum (connected with the frontal brain regions) is linked to impaired memory, automatisation, language (for comparison of verbal IQ in SMAs and DMDs, see Billard et al., 1998) and executive functions (Donders and Taneja, 2009; for general cognitive ability, see Cotton et al., 2001; Banihani et al., 2015; Latimer et al., 2017). This particular genetic pattern is aggravated by negative responses from their environment: DMDs accomplish the motor milestones of the first years of life before gradually experiencing the loss of abilities with increasing feelings of helplessness. The combination of genetic factors that impair motor development and executive cognitive functioning as well as experience-dependent psychological factors result in socio-emotional and behavioral problems (Banihani et al., 2015). Comorbid anxiety, obsessive compulsive disorders as well as Attention Deficit Hyperactivity Disorder and even autism spectrum disorders have been reported. ADHD has been found in about 30% of DMDs, leading to severe learning difficulties and problems in social interaction (Pane et al., 2012). In addition, DMDs often suffer from learning disabilities, such as dyslexia and dyscalculia (Hendriksen and Vles, 2006; for an overview, consult Hendriksen et al., 2011).

Consider now SMAs, where cognitive research supports the conclusion that only their motor ability is selectively compromised. Despite severe impairments in motor development, not even their spatial cognition is affected: Indeed, in a three-location-search-task young SMAs (30 months) made fewer mistakes than healthy and age-matched controls (Rivière and Lécuyer, 2002). SMAs outperformed their healthy peers also in language acquisition (Sieratzki and Woll, 2002; Rivière et al., 2009). von Gontard et al. (2002) investigated cognitive abilities in SMA patients between 6 and 18 years by using different test batteries, excluding subtests with motor components. In contrast to DMDs, they do not experience the loss of their motor capacity. Performance above average in older SMAs (over the age of 11) has been interpreted as a strategy to compensate impaired motor ability through advanced creativity and knowledge that is developing over childhood (von Gontard et al., 2002). SMAs use their high social, emotional and cognitive skills to coordinate care givers towards their individual needs.

Embodied Cognition: The Role of Body-Environment Interactions in Cognition

We now turn to the theoretical motivation for our specific research hypotheses. The recently influential embodied perspective on human cognition rejects the traditional view of the human mind, according to which all our knowledge is mentally represented by amodal symbols contained in semantic memory. On the contrary, it gives the body a crucial role by arguing that cognition uses multi-modal sensory-motor representations originating from the interaction of the organism with its surrounding environment (Barsalou, 1999; Matheson and Barsalou, 2018; for recent reviews, see Fischer and Coello, 2016; Newen et al., 2018). According to embodied cognition, the individual, through action, perception and introspection, acquires concepts which, in turn, are stored in multi-modal memories. These memories are re-activated by simulating selective parts of the acquisition process. As a consequence, even higher-level cognitive processes, such as abstract reasoning or mental arithmetic, rely on involvement of lower-level processes, such as perception and action (e.g., Fischer and Shaki, 2018; Witt, 2018). For example, finger counting is a near-universal mode of acquisition of number concepts that shapes even adult numerical cognition (e.g., Sixtus et al., 2017, 2018).

Following the influential *mental number line* account (Restle, 1970; Dehaene and Cohen, 1995), numbers are represented in the mind as a quantity, based on a spatialized magnitude code. This representation is usually described as continuous, analog, format-independent, and oriented from left to right (in Western cultures; Shaki et al., 2009). Studies on numerical cognition support the hypothesis of a mental number line, revealing the presence of *spatial-numerical associations* (SNAs) and suggesting that spatial processing is crucially involved in numerical cognition. Studies on healthy participants have shown a systematic association between small numbers and left-sided response and between large numbers and right-sided response (Dehaene et al., 1993). Across cultures people represent numbers both on horizontal and vertical mental number lines (Winter et al., 2015).

Embracing the principles of embodied cognition, Fischer (2012) suggests that the spatialization tendency underlying SNAs is not entirely automatic and universal but can be influenced by interactions between the individual and the environment during development. Applying a hierarchical interpretation to SNAs, Fischer (2012) classified them depending on their *grounded aspects* (derived from physical properties present in the real world and universally shared), *embodied aspects* (developed from early experiences of individual-environment interaction and linked to sensory and motor constraints imposed by the body) and *situated aspects* (determined by current contextual constraints). According to this hierarchical view, grounded factors such as the gravity law and the fact that different objects cannot occupy the same space at the same time, explain vertical associations of magnitude, such as MORE IS UP and LESS IS DOWN. Embodied factors such as the habitual use of fingers during counting explain horizontal associations of magnitude with LESS IS LEFT and MORE IS RIGHT, *also* depending on peculiarities of body-context interactions and culture. Finally, situated factors, such as our body postures or temporary characteristics of the surrounding environment, account for the flexibility of the association between space and magnitude (Fischer and Shaki, 2018).

The Current Study

From the embodied cognition perspective, the fact that children affected by severe motor impairment experience different body-context interactions throughout their lives, should result in different cognitive profiles compared to healthy children. Here, we aim to document such impact of limited sensorimotor experiences on cognitive abilities in the specific context of numerical cognition. Numbers exemplify a domain of knowledge that is highly relevant for everyday life, well defined, easy to activate and involving high-level cognitive processes. Consistent with an embodied approach to cognition, sensory and motor associations systematically emerge during the activation of number concepts (Lindemann et al., 2007; Cohen Kadosh et al., 2008; Fabbri et al., 2012). Therefore, lack of motor activity should result in specific cognitive signatures when thinking about numbers.

This prediction led us to examine whether and how sensory and motor constraints present at an early age might influence numerical cognition. To do so, we used the *Random Number Generation* (RNG) task, a cognitive task first developed to assess executive functions (see, Brugger, 1997 for a review and methodological considerations) and more recently become a benchmark test for the investigation of SNAs (Loetscher et al., 2008; Winter and Matlock, 2013; Göbel et al., 2015; Sosson et al., 2018).

Among the first to use RNG in the field of numerical cognition were Loetscher and Brugger (2007). Specifically, they asked healthy adults to randomly generate numbers between 1 and 6 by imagining to throw a die. Several signatures of numerical cognition emerged: First is a small number bias, defined by the tendency to over-generate the first half, i.e. the smaller numbers of the specified range; whereas a large number bias is defined by the over-generation of the second half, i.e. larger numbers of the specified range. The authors found a significant *small number bias*, defined as the tendency to generate more numbers in the range between 1 and 3 rather than in the range between 4 and 6. Later, Towse et al. (2014) explored number processing in children of different age groups with a larger numerical range (from 1 to 10). In contrast to adults, children reported a *large number bias*, defined as the tendency to generate more numbers in the range between 6 and 10 rather than in the range between 1 and 5. Moreover, differences between age groups indicate that the processing of numbers changes in the course of development. As a second signature of numerical cognition, Towse et al. (2014) formulated a first number bias when analyzing number generation data of very young children. This bias is defined as the tendency to over-generate the very first numbers of the specified range. In association with the large number bias, very young children showed an over-generation of numbers 1, 2 and 3, explained by over-representation of these numbers at an early age (Towse et al., 2014).

Loetscher et al. (2008) assessed the influence of body postures by introducing horizontal head movements into the RNG task. Adult participants were asked to randomly produce numbers from 1 to 30 while continuously moving the head along the horizontal axis (from left to right and from right to left). Winter and Matlock (2013) further enriched the paradigm by adding head movements along the vertical axis. The results revealed the presence of numerical biases along both axes: Participants produced larger numbers after moving their head right (up) than after moving their head left (down). Interestingly, this pattern was stronger in the vertical condition, indicative of a more robust vertical mental number line and in line with the hypothesis of a congruent relation between the situated postural influence and the influence of grounded or physical properties of the world on cognition (Fischer, 2012).

More recently, researchers started using the same paradigm to assess SNAs and their developmental changes in healthy children. In light of different SNAs reported in adults depending on culture (Shaki et al., 2009; see Göbel et al., 2011, for a review), Göbel et al. (2015) compared the influence of reading habits of British and Arabic participants on SNAs in adults and children: participants were invited to randomly produce numbers from 1 to 50, while resting either on their right or on their left side. In general, adult participants generated on average larger numbers than children, which is contrary to the established small number bias in adults. The researchers explained this opposite pattern in light of the different numerical range used in the study: indeed, with numbers from 1 to 50, the small number bias is computed on the range from 1 to 25; these numbers might be more strongly represented in children than higher numbers. Importantly, regardless of age group, body position had an effect on the mean of the generated numbers and this effect depended on the participants' habitual reading direction.

While the above studies investigated SNAs in paradigms where numbers were produced while orienting towards a particular side of space, Sosson et al. (2018) compared average of random number generated by adults and children whenever their head reached the central position during horizontally alternating movements. The results revealed significant biases only in adults, highlighting that the amount of sensorimotor experience, or the ability to predict movement outcomes, might influence the strength of SNAs.

To date, SNAs have never been explored in children with neuromuscular diseases. In this study, by using the RNG task we aim to fill this gap in the literature. Moreover, we assess SNAs along both the horizontal and vertical axis in order to explore the role of grounded and embodied factors in number representation of young children with early motor impairment. Young children with DMD reach milestones of early motor development like walking independently (Connolly et al., 2013). Therefore, self-exploration of the environment is equal to healthy children. Following the embodied perspective those early self-exploring experiences are crucial for the development of mental representations (Rakison and Woodward, 2008).

Hypotheses

Inspired by the embodied cognition perspective and the hierarchical interpretation of SNAs (Fischer, 2012) we formulated the following hypotheses on RNG performance and different SNAs in the two groups:

H1: We assume stronger SNAs in DMDs compared to SMAs because of their ability to explore the environment in early childhood and, respectively, weaker SNAs in SMAs compared to DMDs because of their more limited exploration.

H2: Finally, RNG is a suitable task also for assessment of executive functions and their maturation. Given that executive functions are impaired in children affected by DMD (Donders and Taneja, 2009), we expected less random number sequences in children with DMD compared to children with SMA at a given age.

MATERIALS AND METHODS

Participants

Children from 6 to 12 years diagnosed with either SMA (type 1 or type 2) or with DMD were included into the study. In total 16 children participated in the experiment: eight SMA (three type 1 and five type 2; three females and five males) and eight DMD children (all males). Mean age of SMA children was 7.62 years (SD 1.4) and mean age of children with DMD was 9.37 years (SD 1.7), $t_{(14)} = 2.190$; $p = 0.046$. Participants were recruited from the department of pediatrics of the German Red Cross Hospitals Berlin. Diagnosis of SMA and DMD was based on molecular genetic analysis of each participant, provided by the hospital or patients' families. Only children with SMA type 1 and 2 were included because, as mentioned in the Introduction, children with SMA type 3 do not differ from healthy children in milestones of motor development.

The current study was approved by the Ethics Committee of the University of Potsdam (ref. nr. 78/2020).

Tasks and Tests

Wechsler Intelligence Scale for Children (WISC-V)

To assess general cognitive performance, three subtests of the *WISC-V* (German Version, Petermann, 2017) were selected, based on what we wanted to measure (verbal ability, non-verbal cognitive performance and working memory) and the suitability of tests to the population with neuromuscular diseases (indeed, only tests not including motor components were included). Cognitive performance was measured by providing different items in ascending levels of difficulty. To assess verbal ability, the subtest *Similarities* was used. Two terms were given and the participant had to tell in what regard they were alike (e.g., the words *red* and *green*, where the correct answer is that both are colours). For measuring non-verbal cognitive performance, the subtest *Matrix Reasoning* was taken from the battery. Children had to choose from a range of abstract pictures the one that completes a given matrix. Finally, the subtest *Letter-Number-Sequencing* was considered to assess working memory. The participant repeated a given span of mixed numbers and letters by arranging them in ascending and alphabetical order. Age-related items as well as norms were used to assess the individual T-value with normal (age-appropriate) values between $T = 40$ and $T = 60$.

In all participants general cognitive ability was measured using two subtests from the WISC-V, one for verbal ability (Similarities subtest) and one for performance IQ (Matrix Reasoning subtest). Independent sample t-tests revealed no differences between groups, either in the Similarities subtest [$t_{(14)} = 1.479$; $p = 0.161$] or in the Matrix Reasoning subtest [$t_{(14)} = 1.805$; $p = 0.093$]. The test scores (T) of all SMAs in the similarities subtest were in the normal range ($43.3 > T < 60.0$) except for one participant who performed above average ($T = 73.3$). Instead, DMDs' performance ranged from $T = 33.3$ to $T = 60.0$, including two children below average.

Working memory was assessed through the subtest Letter Number Sequencing of the WISC-V. SMAs performed significantly better than DMDs [$t_{(12)} = 6.847$; $p < 0.001$]. Two participants were excluded from the analysis.

Adaptive Intelligence Diagnosticum 3rd Edition (AID 3)

The subtest *Applied Computation* from the *AID 3* (Kubinger and Holocher-Ertl, 2014) was administered for measuring arithmetic skills. Using a branched testing approach, each participant was presented with arithmetic problems that were adapted to their individual level. This ensured high motivation throughout the whole testing. Arithmetic problems included a variety of operations (e.g., additions, subtractions, percentage calculations) in text problem format. An example is: "Petra runs over a 10 m long meadow. Once there and once back. Kurt runs twice as far. How many meters has Kurt covered?" (German: "Petra läuft über eine 10 m lange Wiese. Einmal hin und einmal zurück. Kurt läuft doppelt so weit. Wie viele Meter hat Kurt zurückgelegt?"). Age-related norms were used to assess the individual T-value.

The independent sample t-test comparing both disease groups' arithmetic performance of the Applied Computation subtest of the AID-3 showed that children with SMA performed significantly better (mean T = 55.25) than children with DMD (mean T = 42.75) [$t_{(14)} = 2.426$; $p < 0.05$].

Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern—Revidierte Fassung (ZAREKI-R)

In the subtest *Number Line* from the *ZAREKI-R* (von Aster et al., 2014), 12 trials allowed assessment of spatial representation of numbers along a vertical number line from 1 to 100. The task consists of two parts: In the first part additional markers partition the visual lines into specific portions while in the second part no markers are provided. Participants marked the position of each called-out number on the vertical number line. Age-related norms were used to assess the individual test value.

TABLE 1 | Descriptives of *R* scores in the baseline condition and general cognitive data.

	DMD	SMA	*p* (t-tests)
R Score (baseline)	8.35%	8.47%	n.s. ($p = 0.9$)
WISC-V			
Similarities (verbal IQ)[a]	45.41 (8.90)	52.49 (10.19)	n.s. ($p = 0.1$)
Matrix reasoning (performance IQ)[a]	42.50 (9.36)	52.50 (12.57)	n.s. ($p = 0.09$)
Letter number sequencing (working memory)[a]	30.44 (8.25)	60.57 (8.00)	**$p < 0.001$**
AID-3			
Applied computation[a]	42.75 (7.40)	55.25 (12.56)	**$p < 0.05$**
ZAREKI-R			
Number line (part 1)[b]	89.00 (29.10)	87.29 (28.70)	n.s. ($p = 0.9$)
Number line (part 2)[b]	65.57 (43.41)	73.29 (31.79)	n.s. ($p = 0.7$)

[a]*results presented as t-value with standard deviation.*
[b]*results presented as percentile rank with standard deviation.*
Bold values are indicating significant effects.

The independent sample *t*-test revealed no significant differences between the two groups in both parts of the subtest [part 1: $t_{(12)} = 0.111$; $p = 0.913$; part 2: $t_{(12)} = -0.379$; $p = 0.711$]. Two participants were excluded from the analysis.

Descriptive data of general cognitive abilities are presented in **Table 1**.

Random Number Generation (RNG) Task

Participants generated random numbers between 1 and 10 while moving their head along either the horizontal axis (alternating left-to-right, horizontal condition) or vertical axis (alternating down-to-up, vertical condition) or by keeping the head straight (baseline condition). We considered the following parameters: mean of generated numbers and differences between every generated number and its immediately preceding one, i.e., *first order differences* (FODs). The sign of FODs provides information about direction of movement on the mental number line (positive/negative FODs correspond to ascending/descending tendency), instead, the absolute value of FODs indicates the size of movements on the mental number line. Additionally, a redundancy (*R*) score was computed as an index of randomness (the higher the *R* score, the lower the randomness).

Dot Counting Direction Task (DCD)

During the DCD (Shaki et al., 2012; Fischer and Shaki, 2017), participants sequentially count aloud four dots displayed on a DIN A4 sheet horizontally (left-right) or vertically (up-down) in front of them, by pointing on each dot with their finger. This task is useful to assess directional counting habits. The instructions were: "Please count the number of these dots, by pointing at each of them with your finger" (in German: "Bitte zähle die Punkte nacheinander laut, indem Du mit Deinem Finger auf jeden Punkt zeigst"). We recorded the order of pointing.

Procedure

Due to COVID-19 restrictions, all participants were tested via the online video platform ZOOM. After receiving informed consents from the parents via mail the ZOOM link was sent via mail. To avoid any numerical and lateral cues during the test session, participants were asked to cover the keyboard in front of them and to remove all numerical stimuli (e.g., clocks or pictures) from their surroundings. The experiment was conducted in German language. All participants' parents gave permission to video-record the testing session for further offline evaluations.

Firstly, all participants completed the RNG task. We used a mixed design with Head position (left vs. right vs. up vs. down vs. straight) as within-subjects factor and Disease (SMA vs. DMD) as between-subject factor. For each participant, there were four experimental conditions (left vs. right (horizontal) vs. up vs. down (vertical) as starting position of the head) and one baseline condition (straight). The axes (horizontal vs. vertical vs. straight) were presented in a blocked order while the head positions alternated continuously within each block, along either the horizontal or vertical axis. The experimental conditions (horizontal and vertical) were counterbalanced across participants while the baseline condition was always administered at the end.

Participants were asked to generate numbers between 1 and 10 as randomly as possible while moving their head from left to right (experimental condition horizontal) or up to down (experimental condition vertical). Importantly, the instructions clarified that numbers were only to be stated once the participant's head was in the most extreme (yet still comfortable) position, not before. To facilitate understanding of randomness a ten-sided die was used. A stencil of such a die was sent to the families together with informed consents and the instruction to build it before testing. In our experience, providing participants with this manipulable object facilitates the understanding of the RNG task. The following instruction was used: "Imagine you were a 10-sided-die rolling from one side to another. While rolling please say the sentence "*The die is showing number…* and then please call out a number between 1 and 10." (In German: "Der Würfel zeigt die Nummer..."). Participants were instructed to keep their eyes closed during the complete task. Due to differences in the degree of motor impairment, participants determined their own speed of responding.

After the horizontal and vertical block of the RNG task, the horizontal and vertical version of the DCD was administered. The ZAREKI-R number line subtest followed afterwards. Depending on the fatigue of each participant, assessment of general cognitive abilities was conducted either immediately following or in a separate second session that took place up to 7 days later.

Data Preparation

First of all, omissions, errors (all not numerical oral responses, e.g., letters) and numbers out of the instructed range were detected for each condition and group. Both SMA and DMD groups performed only one omission in the right-side head starting condition and one error in the baseline (head straight) condition. In the SMA group there were four numbers out of the range each in the up and in the down condition, respectively. The DMD group generated numbers out of the range in all conditions: one in the down, three in the up, four in the left, two in the right and three in the baseline condition.

No participant was excluded from the analysis. One participant did not follow the instruction completely and changed the sentence preceding the RNG responses ("*The die is showing number…*") to "*aloha*". Nevertheless, he did not vary this sentence during the trials and his performance did not deviate from that of the others. For these reasons, that participant was also included.

Errors, omissions and numbers out of the range were removed together with their preceding and succeeding numbers (see Sosson et al., 2018, for methodological considerations).

Data Analyses

Firstly, we compared the means of the generated numbers and FODs for each participant individually in all five head positions to detect descriptive group differences followed by statistical analysis using multilevel modelling with the generated number and FODs in each head position as a function of group (DMD vs. SMA). We decided to take that approach to enrich the information gained from the data by emphasizing individual differences. Because of the more fine-grained depiction of the data on the level of head position (left, right, up, down, straight) we can increase power compared to averaging the data as would be done in parametric statistics.

In addition, *R* scores were computed using the software developed by Towse and Neil (1998). To explore differences between the patient groups and normally developing children, baseline *R* scores of DMDs and SMAs were compared to the results reported by Towse and McIachlan (1999) and *R* scores in the horizontal condition were compared to the more recent data by Sosson et al. (2018). Finally, for the two groups, indices for general cognitive ability and arithmetic skills were compared via *t*-tests.

RESULTS

Spatial Numerical Association (Hypothesis 1)

We will start the analysis by taking a closer look at the mean differences between head positions within and between the two groups. Next, we set up multi-level models (random intercepts with responses nested in persons) to test the corresponding hypotheses. In these models we predicted the generated number by head position (model for horizontal positions: left vs. right; model for vertical positions: up vs. down), group (SMA vs. DMD), and the interaction head position × group (see equation 1). Head position and group were dummy coded with left (model 1) and down (model 2) as the reference category for head position and DMD the reference category for group. All estimators were optimized for minimizing the restricted maximum likelihood. Estimation weights are reported non-standardized.

$$\begin{aligned} Generatednumber_{ij} &= \beta 0_j + \beta 1 headposition_{ij} \\ &\quad + \beta 2 headposition_{ij} \times groupj + \epsilon_{ij}. \\ \beta 0_j &= \gamma 00 + \gamma 01 group_{1j} + u0_j \\ i &:= trial \\ j &:= person \end{aligned} \quad (1)$$

The analyses were conducted with the R package lme4 (Bates et al., 2015).

Additionally, FODs were taken instead of the magnitude of the generated number as the dependent variable, we also conducted all analyses and models with FODs. The results are nearly identical and will be reported in the Tables 5, 6 in **Supplementary Material** of this paper to avoid redundancy.

Description of the Means

Following Sosson et al. (2018), we start by reporting the mean of generated numbers and FODs, separate for all head positions (left, right, down, up, straight) and groups (SMA, DMD), present in **Table 2**.

In the horizontal condition the mean of generated numbers is higher in children with SMA than DMD. Additionally, in children with DMD the mean of generated numbers is higher in the right-head (M = 5.99; SD = 2.79) than in the left-head position (M = 5.80; SD = 2.94), whereas in children with SMA it is similar in the left (M = 5.53; SD = 2.98)- and right-head position (M= 5.54; SD = 3.04). In the vertical condition children with SMA produced fewer small numbers when turning the head down (M = 6.1; SD = 2.83) compared to up (M = 5.54; SD = 3.07), whereas children with DMD reported the reversed pattern. They produced more small numbers in the down (M = 5.28; SD = 2.83)- and larger numbers in the up-head position (M = 5.57; SD = 3.04).

The following pattern was found in FODs: In the horizontal condition, children with SMA produced more ascending sequences when turning their head left (M = 0.21; SD = 4.41) than right (M = −0.10; SD = 3.49), while the opposite pattern was observed in DMD children with more ascending sequences generated during right (M = 0.08; SD =3.86) rather than left (M = −0.07; SD = 3.49) head position. In the vertical condition children with SMA tended to generate more ascending sequences (represented by positive FODs) when turning the head down (M = 0.69; SD = 3.82) rather than up (M = −0.75; SD = 4.36). A reversed pattern was observed in DMD children, who performed more ascending steps when turning the head down (M = −0.33; SD = 3.80) rather than up (M = 0.27; SD = 3.72). The FODs measures corroborated tendencies revealed by the previously considered mean of generated numbers.

Hypothesis Testing

Table 3 shows the estimations for the multi-level model of the horizontal condition. Firstly, we included *head position* as a predictor (model 1) and then the *group and head position × group* (model 2). Neither *head position* (B = 011, p = 0.616), *group* (B = −0.26, p = 0.450) nor the interaction *head position × group* (B

TABLE 2 | Descriptive information for all head positions in both groups.

Group	Head position	Mean number (sd)	Mean FOD (sd)	Total of omissions	Total of errors	Total *n* out
DMD	Left	5.80 (2.94)	−0.07 (3.49)	0	0	4
	Right	5.99 (2.79)	0.08 (3.86)	1	0	2
	Down	5.28 (2.83)	−0.33 (3.80)	0	0	1
	Up	5.57 (3.04)	0.27 (3.72)	0	0	3
	Straight	5.49 (2.87)	0.05 (3.62)	0	1	3
SMA	Left	5.53 (2.98)	0.21 (4.14)	0	0	0
	Right	5.54 (3.04)	−0.10 (4.19)	1	0	0
	Down	6.10 (2.83)	0.69 (3.82)	0	0	4
	Up	5.42 (3.07)	−0.75 (4.36)	0	0	4
	Straight	5.61 (3.02)	0.09 (4.32)	0	1	0

TABLE 3 | Multilevel model for mean of generated numbers in left-right head movements (horizontal).

	Model 1				Model 2			
Predictors	***B***	***se***	***t***	***p***	***B***	***se***	***t***	***p***
(Intercept)	5.66	0.18	31.94	**<0.001**	5.80	0.25	23.02	**<0.001**
Head position (right)	0.11	0.23	0.50	0.616	0.20	0.33	0.60	0.545
Group (SMA)					−0.26	0.35	−0.76	0.450
Head position (right) × Group (SMA)					−0.19	0.46	−0.40	0.686
Random effects								
σ^2	8.56				8.57			
τ_{00}	0.07 $_{ID}$				0.05 $_{ID}$			
ICC	0.01				0.01			
N	16 $_{ID}$				16 $_{ID}$			
Observations	657				657			
Marginal R^2/Conditional R^2	0.000/0.008				0.004/0.010			

Bold values are indicating significant effects.

$= -0.19$, $p = 0.686$) significantly impacted the magnitude of the generated numbers.

In the vertical condition (up-down head movements, **Table 4**) we, again, did not find a significant effect of *head position* (B $= -0.18$, $p = 0.435$). That is, *head position* had no impact on the generated number for the DMD group. But the generated numbers were significantly higher for the *SMA group* (B $=$ 0.83, $p < 0.05$) and the *SMA group* also had a significantly stronger effect of the *head position* compared to the DMD group (B $= -0.97$, $p < 0.05$).

Randomness Performance (Hypothesis 2)

The redundancy scores (R) in both groups were computed separately for both head positions along the vertical and horizontal axis and the baseline (straight head position) by using the software from Towse and Neil (1998). R scores can range between 0 and 100% where 0% means complete randomness in number generation and 100% results from a completely predictable sequence. The observed average R scores were 8.47% (range 1.14–16.66%) for SMA children and 8.35% (range 1.78–17.47%) for DMD children. A two-tailed independent t-test limited to the straight head condition showed no significant difference between SMAs and DMD's R scores [$t_{(14)} = -0.042$; $p = 0.967$, **Table 1**].

Next, we compared the performance of our sample against that of healthy children by referring to published R values in comparable conditions. Towse and McIachlan (1999) assessed the RNG task with children of different age groups by keeping the head straight. Two-tailed single sample t-tests for each group in the straight head condition were computed against the reference value. We compared the R values of the SMA group against $R = 2.29\%$, a value that has been reported for a sample of approximately 8 years olds. We obtained a significant deviation from those reference values of healthy children in SMAs [$t_{(7)} = 3.082$; $p = 0.018$; $R = 8.47\%$]. The score of $R = 8.35$ % for our DMDs was compared against the reference value of $R = 4.18$ % of the older subsample (average age 10 years). No significant group difference was revealed [$t_{(7)} = 1.899$; $p = 0.099$; $R = 8.35\%$]. Sosson et al. (2018) introduced left/right head movements to the RNG task. We compared performance of our two groups in the horizontal condition with their R score of 10.64% for healthy children (age range 7.8–11.9 years). A two-tailed single sample t-test for each group in the horizontal condition was computed. t-tests revealed significant deviations from reference

TABLE 4 | Multilevel model for the generated numbers in up-down head movements (vertical).

	Model 1				Model 2			
Predictors	***B***	***se***	***t***	***p***	***B***	***se***	***t***	***P***
(Intercept)	5.67	0.18	31.99	**<0.001**	5.28	0.24	21.66	**<0.001**
Head position (up)	−0.18	0.23	−0.78	0.435	0.29	0.33	0.89	0.374
Group (SMA)					0.83	0.35	2.36	**0.018**
Head position (up) × Group (SMA)					−0.97	0.47	−2.08	**0.037**
Random effects								
σ^2	8.65				8.60			
τ_{00}	0.07 $_{ID}$				0.06 $_{ID}$			
ICC	0.01				0.01			
N	16 $_{ID}$				16 $_{ID}$			
Observations	636				636			
Marginal R^2/Conditional R^2	0.001/0.009				0.011/0.018			

Bold values are indicating significant effects.

values of healthy children in SMAs [$t_{(7)} = -7.128$; $p<0.001$; $R = 3.65\%$] as well as in DMDs [$t_{(7)} = -5.650$; $p < 0.005$; $R = 4.49\%$].

Exploratory Analysis: Dot Counting Task

All participants counted the four dots successively from one side to the other. All SMAs counted from left to right except for one participant who counted from right to left. Interestingly, this child also generated on average larger random numbers with left-turned head (RNG $M = 6.27$) than with right-turned head (RNG $M = 5.67$). Three of eight DMD participants counted from right to left (left-turned head: RNG $M = 5.50$; right-turned head: RNG $M = 5.59$).

On the vertical axis counting habits were the following: For SMAs again only one participant counted from down (RNG $M = 6.00$) to up (RNG $M = 5.60$). Among DMDs two of eight children counted from down to up (down-turned head: RNG $M = 5.31$; up-turned head: RNG $R = 6.55$).

DISCUSSION

Inspired by an embodied approach to cognition, the current study aims to investigate number processing and higher order cognitive abilities in children with rare neuromuscular diseases. We tested eight SMA and eight DMD children between 6 and 12 years in a Random Number Generation task. Since it is a suitable task for numerical cognition as well as for executive functions, we derived two distinct predictions. Below, we discuss the findings following our hypotheses: Hypothesis 1 is related to spatial numerical associations, hypothesis 2 is related to randomness, indicative of inhibitory control.

Spatial-Numerical Associations (Hypothesis 1)

For the first time, spatial-numerical associations (SNAs) were assessed in children with motor impairments, examining their SNAs along both the horizontal and the vertical axis. We hypothesized, that SNAs would differ between children with SMA and DMD due to differences in the extent of self-exploration in early childhood. Although patterns of SNAs in children with SMA were atypical, they performed above average in general cognitive and arithmetic tasks. Participants with DMD, on the other hand, revealed SNAs comparable to those expected of healthy children but showed weaker performance in arithmetic and in general cognitive ability and executive functioning.

Even though no systematic difference in the counting direction was found in the DCD between DMDs and SMAs, participants with DMD had a tendency to generate larger/smaller numbers after turning their head rightward/leftward, but the ones with SMA did not present a preference in the left-right direction. This may perhaps be related to the severe and early handicap in changing their postural positions including an inability to take an upright body posture as well as to a less mature development of functional laterality in the significantly younger SMAs.

Individual differences within as well as between the two groups were observed for horizontal SNAs and can be interpreted following the hierarchical view on embodied cognition (Fischer, 2012). The differences between the groups can be explained by distinct motor experiences in the course of the two diseases, i.e., on the embodied level. Additionally, in both groups for some participants, testing was only possible with a caregiver next to them. The lateral presence of another person might anchor the participant and therefore has an impact on number generation, i.e., on the situational level. Three out of seven participants with a caregiver sitting next to them did show reversed or no horizontal SNAs.

Concerning next the vertical dimension, DMDs showed patterns comparable to healthy developing children by performing more descending/ascending steps in the downward/upward head orientation, respectively. In contrast, SMAs revealed a reversed tendency. How to explain this opposite trend? From an embodied point of view, early sensorimotor experience is crucial for developing mental representations. First, disease onset in DMD is around age 6–7 which indicates that several milestones of motor development are reached before outbreak of the disease and loss of muscle strength. In contrast, SMA children do not typically acquire motor skills such as sitting or standing autonomously. Therefore, compared to SMAs, experiences of DMDs do not differ that much from healthy developing children in those first and important years of life. Their results in this task of embodied number processing can therefore be equal to the ones of healthy children.

An additional interpretation of the unique pattern of RNG performance in children with SMA along the vertical axis pertains to a different internalization of universal concepts related to physical properties of the world. Indeed, due to their inability to reach low or high points in space, SMAs could be more sensitive to an object's weight. From an early age, they start experiencing the weight of their limbs and as they grow and become taller, their body gets heavier without gaining more muscle strength like healthy children. Associations might be built between more weight (i.e., heavy) and downward space and less weight (i.e., light) and upward space, respectively. The vertical spatial representation of weight is a rarely studied but upcoming topic. Across domains "more" is associated with "up" and "less" is associated with "down". According to the hierarchical view on embodied cognition, weight could play a distinct role on the situated level: A recent study by Vicovaro and Dalmaso (2020) found a reversed SNARC-like effect when participants were directly instructed towards the weight of objects. They responded faster with a downward button for higher weights and with an upward button for lower weights. This effect was only found when weight was task relevant. This finding could support our interpretation: The weight of an object is of specific valence for children with SMA.

RNG and Randomness (Hypothesis 2)

We measured the randomness of our participants with SMA and DMD by computing the redundancy score *R*. Low redundancy equals high randomness in number generation. Impairments in executive functions are linked to the inability to inhibit certain tendencies in number production (see Brugger, 1997; Peters et al., 2007). Research with DMD children revealed impairment in executive functions (Donders and Taneja, 2009). Therefore, we hypothesized that performance of children with DMD would have been less random than performance of participants with SMA. However, both groups' redundancy did not differ significantly in the current study. Developmental improvements in performance were established in studies on random generation with children without motor impairment (Towse and McIachlan, 1999). Additionally, *R* scores of both groups differed significantly from *R* scores of healthy children on the horizontal axis (Sosson et al., 2018) but only for SMAs when excluding the spatial component from the task (Towse and McIachlan, 1999). It is worth noticing that, since in Sosson et al. (2018) the numerical range was larger than the one used in our and in the Towse and McIachlan (1999) study and since increase of numerical range reduces the randomness (Towse and McIachlan, 1999), the comparison with the index computed by Sosson et al. (2018) should be considered with precautions. Nevertheless, higher age in DMDs contributes to them performing more similar to healthy controls as executive functions as well as motor experience evolve with age.

Cognitive Profile

Since SMA and DMD are associated with several differences in general cognitive abilities, we assessed the cognitive profile of the two groups by administering subtests from established test batteries (WISC-V and AID-3). Verbal and non-verbal performance was assessed by two subtests from the WISC-V. It was expected that children with SMA would outperform children with DMD in verbal IQ whereas performance related assessment would not show any differences. The current study did not find significant distinctions in verbal ability between the two groups, even though tendentially SMA children did outperform DMD boys. The lack of a significant difference in the predicted direction might also reflect the reliable age advantage of DMD children in our study.

Limitations and Outlook

The current study fell right into the COVID19-crisis. Testing participants in conditions affecting the respiratory system made avoidance of any infection risk even more important. We decided to test the participants online via the platform ZOOM. Despite the advantage of being more flexible in scheduling the appointments and incurring less effort for the families, some limitations had to be accepted.

In light of contextual influences on number processing (Fischer et al., 2010), RNG performance may be sensitive to task-irrelevant numerical cues in participants' visual field. In our online study, it was not possible to sufficiently control for such visual cues, as well as for auditory noise coming from other persons or even construction work. Setting up and monitoring the technical equipment required care givers being close to the participants. Exploring the influence of lateralized cues on RNG performed by SMAs and DMDs would identify the weights of contextual factors (Fischer, 2012).

One potential limitation of the current study is the missing age matched control group of healthy children. Here, we compared performance of children with rare neuromuscular diseases to reported results in former studies with healthy and normally developing children by Towse and McIachlan (1999) and Sosson et al. (2018). Future research should provide a direct comparison of children with SMA and DMD with age matched healthy controls.

Finally, the rarity of SMA and DMD led to rather small samples and contributed to the significant age difference between the two groups. Bigger sample sizes in future research should compare groups with different degrees of motor impairment and finger dexterity (Guedin et al., 2018).

CONCLUSION

To summarize, DMDs, unlike SMAs, seem to have both a horizontal and a vertical mental number line, similar to healthy children. In other words, DMDs' performance supports the hypothesis that, due to equal sensorimotor experience, the mental representation of numbers of DMDs is comparable to that of healthy developing children. The observed discrepancy between typically developing SNAs and atypically (weak) developing math abilities in DMDs is likely due to deficits in the development of supportive domain general cognitive functions, like executive and attentional regulation, as well as due to psychological coping difficulties with gradual loss of motor ability. For SMAs it seems unlikely that the observed atypical SNAs are at all disadvantageous for the development of their high math abilities. They seem to be able to spatially mentalize numbers albeit their SNAs develop atypically in spatial direction and strength.

In general, this supports the view that typical as well as atypical development of numerical cognition cannot be predicted by single factor models (i.e., core ability / deficit) but rather by multiple factor models that cover a wide range of biological as well as environmental individual differences (Kaufmann et al., 2013). Moreover, our results support and further develop the hierarchical view by Fischer (2012), demonstrating how the properties of the body inhabiting the brain and the development of compensatory skills are determinant in the use of different spatial information (derived from physical properties or acquired through daily activities) for number representation and processing.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the University of Potsdam (ref. nr. 78/2020). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

HS, AF, and MF: conceived and designed the experiments. HS: collected data. HS, AF, and JW: analyzed the data. HS and AF: wrote the paper. AM: advised in medical terminology and provided contact to participants. JW, MA, and MF: advised in methodological aspects. MF, AM, JW, and MA: proofread the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Linda Kuhr for advises during the process of participant recruiting.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.697881/full#supplementary-material

REFERENCES

Bachot, J., Gevers, W., Fias, W., and Roeyers, H. (2005). Number sense in children with visuospatial disabilities: orientation of the mental number line. *Psychol. Sci.* 47, 172–183.

Banihani, R., Smile, S., Yoon, G., Dupuis, A., Mosleh, M., Snider, A., et al. (2015). Cognitive and neurobehavioral profile in boys with Duchenne muscular dystrophy. *J. Child Neurol.* 30, 1472–1482. doi: 10.1177/0883073815570154

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Billard, C., Gillet, P., Barthez, M. A., Hommet, C., and Bertrand, P. (1998). Reading ability and processing in Duchenne muscular dystrophy and spinal muscular atrophy. *Dev. Med. Child Neurol.* 40, 12–20. doi: 10.1111/j.1469-8749.1998.tb15351.x

Billard, C., Gillet, P., Signoret, J. L., Uicaut, E., Bertrand, P., Fardeau, M., et al. (1992). Cognitive functions in Duchenne muscular dystrophy: a reappraisal and comparison with spinal muscular atrophy. *Neuromuscul. Disord.* 2, 371–378. doi: 10.1016/S0960-8966(06)80008-8

Brugger, P. (1997). Variables that influence the generation of random sequences: an update. *Percept. Mot. Skills* 84, 627–661. doi: 10.2466/pms.1997.84.2.627

Butterworth, B. (2018). The implications for education of an innate numerosity-processing mechanism. *Philos. Trans. R. Soc. B: Biol. Sci.* 373:20170118. doi: 10.1098/rstb.2017.0118

Calucho, M., Bernal, S., Alías, L., March, F., Venceslá, A., Rodríguez-Álvarez, F. J., et al. (2018). Correlation between SMA type and SMN2 copy number revisited: an analysis of 625 unrelated Spanish patients and a compilation of 2834 reported cases. *Neuromuscul. Disord.* 28, 208–215. doi: 10.1016/j.nmd.2018.01.003

Cohen Kadosh, R., Cohen Kadosh, K., and Henik, A. (2008). When brightness counts: the neuronal correlate of numerical–luminance interference. *Cerebral Cortex* 18, 337–343. doi: 10.1093/cercor/bhm058

Connolly, A. M., Florence, J. M., Cradock, M. M., Malkus, E. C., Schierbecker, J. R., Siener, C. A., et al. (2013). Motor and cognitive assessment of infants and young boys with Duchenne Muscular Dystrophy: results from the Muscular Dystrophy Association DMD Clinical Research Network. *Neuromuscul. Disord.* 23, 529–539. doi: 10.1016/j.nmd.2013.04.005

Cotton, S., Voudouris, N. J., and Greenwood, K. M. (2001). Intelligence and Duchenne muscular dystrophy: full-scale, verbal, and performance intelligence quotients. *Dev. Med. Child Neurol.* 43, 497–501. doi: 10.1017/S0012162201000913

D'Angelo, M. G., and Bresolin, N. (2006). Cognitive impairment in neuromuscular disorders. *Muscle Nerve* 34, 16–33. doi: 10.1002/mus.20535

Deenen, J. C., Horlings, C. G., Verschuuren, J. J., Verbeek, A. L., and van Engelen, B. G. (2015). The epidemiology of neuromuscular disorders: a comprehensive overview of the literature. *J. Neuromuscl. Dis.* 2, 73–85. doi: 10.3233/JND-140045

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

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

Donders, J., and Taneja, C. (2009). Neurobehavioral characteristics of children with Duchenne muscular dystrophy. *Child Neuropsychol.* 15, 295–304. doi: 10.1080/09297040802665777

Fabbri, M., Cancellieri, J., and Natale, V. (2012). The a theory of magnitude (ATOM) model in temporal perception and reproduction tasks. *Acta Psychol.* 139, 111–123. doi: 10.1016/j.actpsy.2011.09.006

Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cogn. Process.* 13, 161–164. doi: 10.1007/s10339-012-0477-5

Fischer, M. H., and Coello, Y. E. (2016). *Foundations of Embodied Cognition: Conceptual and Interactive Embodiment*. London: Routledge/Taylor and Francis Group.

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. (2017). Implicit spatial-numerical associations: negative numbers and the role of counting direction. *J. Exp. Psychol. Human Percep. Perform.* 43:639. doi: 10.1037/xhp0000369

Fischer, M. H., and Shaki, S. (2018). Number concepts: abstract and embodied. *Philos. Trans. R. Soc. B: Biol. Sci.* 373:20170125. doi: 10.1098/rstb.2017.0125

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

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

Guedin, N., Fluss, J., and Thevenot, C. (2018). Dexterity and finger sense: a possible dissociation in children with cerebral palsy. *Percept. Mot. Skills*. 125, 718–731.

Hendriksen, J., Hendriksen, R., Kuijer, J., and Vroom, E. (2011). The psychology of Duchenne muscular dystrophy. *J. Hendriksen & Parent Project Production*, ISBN/EAN, 978–90.

Hendriksen, J. G., and Vles, J. S. (2006). Are males with Duchenne muscular dystrophy at risk for reading disabilities?. *Pediatr. Neurol.* 34, 296–300. doi: 10.1016/j.pediatrneurol.2005.08.029

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

Kolb, S. J., and Kissel, J. T. (2015). Spinal muscular atrophy. *Neurol. Clin.* 33, 831–846. doi: 10.1016/j.ncl.2015.07.004

Kubinger, K. D., and Holocher-Ertl, S. (2014). *AID 3: Adaptives Intelligenz Diagnostikum 3*. Göttingen: Hogrefe.

Latimer, R., Street, N., Conway, K. C., James, K., Cunniff, C., Oleszek, J., et al. (2017). Secondary conditions among males with Duchenne or Becker muscular dystrophy. *Journal of Child Neurology* 32, 663–670. doi: 10.1177/0883073817701368

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80, 155–165. doi: 10.1016/0092-8674(95)90460-3

Lindemann, O., Abolafia, J. M., Girardi, G., and Bekkering, H. (2007). Getting a grip on numbers: numerical magnitude priming in object grasping. *J. Exp. Psychol. Human Percept. Perform.* 33:1400. doi: 10.1037/0096-1523.33.6.1400

Loetscher, T., and Brugger, P. (2007). Exploring number space by random digit generation. *Exp. Brain Res.* 180, 655–665. doi: 10.1007/s00221-007-0889-0

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

Matheson, H. E., and Barsalou, L. W. (2018). "Embodiment and grounding in cognitive neuroscience," in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, eds J. Wixted (Hoboken: John Wiley and Sons), 1–27. doi: 10.1002/9781119170174.epcn310

Newen, A., Gallagher, S., and De Bruin, L. (2018). "4E Cognition," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, S. Gallagher and L. De Bruin (Oxford: Oxford University Press), 3–15. doi: 10.1093/oxfordhb/9780198735410.001.0001

Pane, M., Lombardo, M. E., Alfieri, P., D'Amico, A., Bianco, F., Vasco, G., et al. (2012). Attention deficit hyperactivity disorder and cognitive function in Duchenne muscular dystrophy: phenotype-genotype correlation. *J. Pediatr.* 161, 705–709. doi: 10.1016/j.jpeds.2012.03.020

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

Petermann, F. (2017). *WISC-V: Wechsler Intelligence Scale for Children—deutsche* Bearbeitung. London: Pearson.

Peters, M., Giesbrecht, T., Jelicic, M., Merckelbach, H., American Psychiatric Association, H., Anastasi, A., et al. (2007). The random number generation task: psychometric properties and normative data of an executive function task in a mixed sample. *J. Int. Neuropsychol. Soc.* 13, 626–634. doi: 10.1017/S1355617707070786

Rakison, D. H., and Woodward, A. L. (2008). New perspectives on the effects of action on perceptual and cognitive development. *Dev. Psychol.* 44:1209. doi: 10.1037/a0012999

Restle, F. (1970). Speed of adding and comparing numbers. *J. Exp. Psychol.* 83, 274–278. doi: 10.1037/h0028573

Rivière, J., and Lécuyer, R. (2002). Spatial cognition in young children with spinal muscular atrophy. *Dev. Neuropsychol.* 21, 273–283. doi: 10.1207/S15326942DN2103_4

Rivière, J., Lécuyer, R., and Hickmann, M. (2009). Early locomotion and the development of spatial language: Evidence from young children with motor impairments. *Eur. J. Dev. Psychol.* 6, 548–566. doi: 10.1080/17405620701345712

Schaaf, C., and Zschocke, J. (2018). *Basiswissen Humangenetik*. Berlin: Springer-Verlag. doi: 10.1007/978-3-662-56147-8

Shaki, S., Fischer, M. H., and Göbel, S. M. (2012). Direction counts: a comparative study of spatially directional counting biases in cultures with different reading directions. *J. Exp. Child Psychol.* 112, 275–281. doi: 10.1016/j.jecp.2011.12.005

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

Sieratzki, J. S., and Woll, B. (2002). Toddling into language: precocious language development in motor-impaired children with spinal muscular atrophy. *Lingua*. 112, 423–433.

Sixtus, E., Fischer, M. H., and Lindemann, O. (2017). Finger posing primes number comprehension. *Cogn. Process.* 18, 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

Spuler, S., and von Moers, A. M. (2004). *Muskelkrankeiten: Grundlagen, Diagnostik und Therapie.* Stuttgart: Schattauer.

Taylor, P. J., Betts, G. A., Maroulis, S., Gilissen, C., Pedersen, R. L., Mowat, D. R., et al. (2010). Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. *PloS ONE* 5:e8803. doi: 10.1371/journal.pone.0008803

Towse, J. N., Loetscher, T., and Brugger, P. (2014). Not all numbers are equal: preferences and biases among children and adults when generating random sequences. *Front. Psychol.* 5:19. doi: 10.3389/fpsyg.2014.00019

Towse, J. N., and McIachlan, A. (1999). An exploration of random generation among children. *Br. J. Dev. Psychol.* 17, 363–380. doi: 10.1348/026151099165348

Towse, J. N., and Neil, D. (1998). Analyzing human random generation behavior: a review of methods used and a computer program for describing performance. *Behav. Res. Methods Instrum. Comput.* 30, 583–591. doi: 10.3758/BF03209475

Vicovaro, M., and Dalmaso, M. (2020). Is 'heavy' up or down? Testing the vertical spatial representation of weight. *Psychol. Res.* 85, 1183–1200. doi: 10.1007/s00426-020-01309-0

von Aster, M. G., Kaufmann, L., McCaskey, U., and Kucian, K. (2021). "Rechenstörungen im Kindesalter," in *Psychiatrie und Psychotherapie* *des Kindes- und Jugendalters*, eds J. Fegert et al. Heidelberg: Springer Reference Medizin.

von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

von Aster, M. G., Weinhold-Zulauf, and Horn. (2014). *Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern – Revidierte Fassung* (ZAREKI-R).

von Gontard, A., Zerres, K., Backes, M., Laufersweiler-Plass, C., Wendland, C., Melchers, P., et al. (2002). Intelligence and cognitive function in children and adolescents with spinal muscular atrophy. *Neuromuscul. Disord.* 12, 130–136. doi: 10.1016/S0960-8966(01)00274-7

Winter, B., and Matlock, T. (2013). More is up and right: Random number generation along two axes. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35. Retrieved from https://escholarship.org/uc/item/3cx722ps

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

Witt, J. K. (2018). "Perception and action," in *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, vol 2,* eds J. Wixted (Hoboken: John Wiley and Sons), 1–35. doi: 10.1002/9781119170174.epcn211

Woloshin, S., Schwartz, L. M., Moncur, M., Gabriel, S., and Tosteson, A. N. (2001). Assessing values for health: numeracy matters. *Med. Decis. Mak.* 21, 382–390. doi: 10.1177/0272989X0102100505

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Education

ORIGINAL RESEARCH
published: 14 September 2021
doi: 10.3389/feduc.2021.683424

Check for updates

The First Step to Learning Place Value: A Role for Physical Models?

Lei Yuan[1]*, *Richard Prather*[2], *Kelly Mix*[2] *and Linda Smith*[3,4]

[1]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, United States, [2]Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, United States, [3]Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, United States, [4]School of Psychology, University of East Anglia, Norwich, United Kingdom

Very few questions have cast such an enduring effect in cognitive science as the question of "symbol-grounding": Do human-invented symbol systems have to be grounded to physical objects to gain meanings? This question has strongly influenced research and practice in education involving the use of physical models and manipulatives. However, the evidence on the effectiveness of physical models is mixed. We suggest that rethinking physical models in terms of analogies, rather than groundings, offers useful insights. Three experiments with 4- to 6-year-old children showed that they can learn about how written multi-digit numbers are named and how they are used to represent relative magnitudes based on exposure to either a few pairs of written multi-digit numbers and their corresponding names, or exposure to multi-digit number names and their corresponding physical models made up by simple shapes (e.g., big-medium-small discs); but they failed to learn with traditional mathematical manipulatives (i.e., base-10 blocks, abacus) that provide a more complete grounding of the base-10 principles. These findings have implications for place value instruction in schools and for the determination of principles to guide the use of physical models.

Keywords: symbol-grounding, relational mapping, place value, number, analogy, symbol systems

Edited by:
Korbinian Moeller, Loughborough University, United Kingdom

Reviewed by:
Brian Rivera, University of Alabama, United States
Martin H. Fischer, University of Potsdam, Germany
Hendrikje Schmidt, University of Potsdam, Germany, in collaboration with reviewer MF

***Correspondence:**
Lei Yuan
Lei.Yuan@colorado.edu

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

Received: *21 March 2021*
Accepted: *27 August 2021*
Published: *14 September 2021*

Citation:
Yuan L, Prather R, Mix K and Smith L (2021) The First Step to Learning Place Value: A Role for Physical Models?. Front. Educ. 6:683424. doi: 10.3389/feduc.2021.683424

INTRODUCTION

Do symbols need to be grounded to their physical referents to have meanings? Thirty years ago, Harnad (1990) posed this question motivating empirical studies and theoretical debates across many subfields of cognitive science (Taddeo and Floridi, 2005; Steels, 2008; De Vega et al., 2008; Dove, 2016). The issues are still theoretically (Socher et al., 2013; Wang et al., 2019) and practically relevant, especially within the field of education (Alibali and Nathan, 2012; Pouw et al., 2014; Stolz, 2015). How do symbols—the letters of the alphabet, the digits of Arabic numbers—become able to convey meaning? A grounded symbol is one that gains meaning directly through the perception of that meaning, as the symbol "7" may gain meaning through the direct perception of seven discrete entities. In mathematics education, theoretical ideas about symbol grounding influenced and encouraged the use of physical models and manipulatives as a way to make abstract concepts directly perceivable (Sowell, 1989; Sarama and Clements, 2009; Carbonneau et al., 2013). Efficacy studies, however, yielded mixed results (Son et al., 2008; McNeil et al., 2009; Carbonneau et al., 2013; Mix et al., 2014; Mix et al., 2017) and no clear principles as to when physical models are helpful. Here, we propose a rethinking of physical models in education—not as a path to grounding, but as analogies that help learners discover inherently abstract relations. We consider these ideas with respect to children's early learning about multi-digit notation.

Multi-digit Arabic numbers represent magnitudes through a base-10 system: 232 is named as "two hundred thirty-two", and is composed of 2 sets of 100 s, 3 sets of 10 s, and 2 sets of 1 s. 232 is less than 322, because the former indicates 2 rather than 3 sets of hundreds—even though these two numbers are composed of the exact same set of individual digits. Understanding this hierarchical structure and the algebraic relations within multi-digit numbers is the goal of formal place value instruction and the foundation for developing advanced calculation skills. Both research and educational practice (Montessori, 1917; Bruner, 1966; Fuson, 1986; Fuson and Briars, 1990; Geary, 2007; Bussi, 2011) have focused on how to ground base-10 relations (e.g., 100 is 10 sets of 10, and 10 is 10 sets of 1) in physical models, sometimes also known as mathematical manipulatives. For example, base-10 blocks ground the meanings of the counts of each place, the multiplicative relations among the places, and the exact represented discrete quantities using blocks composed of small cubes such that each cube represents one, bars (called "longs" in math education) contain 10 ones, and large blocks (called "flats" in math education) contain 100 ones. These blocks are then used to physically instantiate specific amounts such that "232" is represented as 2 big blocks, 3 bars, and 2 cubes. Despite widespread use, early learners can have difficulties in understanding just what this is all about: some when shown a display such as that for "232" count the total number of blocks (seven blocks), some try to count all the cubes (Chan et al., 2014), and some studies show little benefit of the addition of these physical models (Ball, 1992). We believe that the problem may be that these blocks try to provide a full grounding of the unit size of places, their counts, and the exact quantity, and in doing so, base-10 blocks—just like the base-10 system itself—are too much for a naïve learner to grasp all at once. A complete grounding of the base-10 hierarchy may benefit later learning, but it may not be the best way to introduce the multi-digit number system.

Recent studies suggest that children's learning about place value starts early and proceeds incrementally (Byrge et al., 2014; Mix et al., 2014; Mix et al., Under review; Yuan et al., 2019). Children first learn about the place relations that structure multi-digit numbers before formal school instruction on place value. This early understanding does not include precise knowledge of the different quantities represented by the places nor their multiplicative relations to each other (Byrge et al., 2014; Mix et al., 2014; Yuan et al., 2019; Yuan et al., 2020). Instead, it is an "approximate" understanding that multi-digit numbers are made up of places that represent different relative magnitudes ordered from left to right. Critically, this early approximate knowledge strongly predicts later success in learning and using base-10 principles (Mix et al., Under review), suggesting that approximate understanding is a useful step to more explicit correct understanding. Current evidence suggests further that young children acquire this approximate understanding through experience with the correspondences between spoken and written number names, e.g., learning that the "hundred" and the "-ty" in "two hundred thirty two" mark the places in the string "232" as signifying different amounts (Mix et al., Under review; Yuan et al., 2020). Multiple experiments (Byrge et al., 2014; Mix et al., 2014; Mix et al., 2017; Yuan et al., 2019) have shown that this partial knowledge enables preschool and kindergarten children to map unfamiliar number names to written forms for 3- and 4-digit numbers, to judge the relative magnitudes of 3- and 4-digit numbers, and to write multi-digit numbers given the spoken name (albeit sometimes with meaningful and interesting errors) (Mix et al., 2014). The three experiments reported here focus on this first step in learning about place value: that there are different places that signify different amounts. We ask whether and how physical models might benefit this learning.

In science, physical models are often used not to ground meaning but as analogies to distill the skeleton of an idea: for example, an atom is like the Solar System in that each has smaller elements rotating around a larger one (Gentner, 1983). These simple analogies are helpful to initialize learning and can support higher conceptual inferences, but they are not fully correct (Gentner and Stevens, 1983; Mix, 2010; Mix et al., 2019; Richland et al., 2017; Richland and Simms, 2015). Gentner's Structure Mapping theory (Gentner, 1983; Gentner, 2010) proposes that analogies work because they support the alignment of two relational systems that enable the relations—independent of the elements in those relations—to be extracted. The key to extracting the common relational structure is the alignment and mapping of corresponding elements—e.g., the nucleus to the Sun, the planets to electrons. Experiments with many kinds of materials and different aged participants show that relational structures can be discovered and broadly *generalized* in very few trials if the learner properly aligns the elements across examples (Loewenstein and Gentner, 2001; Namy and Gentner, 2002; Rattermann and Gentner, 1998; Yuan et al., 2017). **Figure 1** provides an illustration of the relevant findings. Given an array such as that in **Figure 1A**, 4-year-olds do not immediately see the big-medium-small structure and have considerable difficulty at picking out another configuration that exemplifies the same relational structure (Kotovsky and Gentner, 1996). But adding another configuration as shown in **Figure 1B** and inviting children to compare the two significantly increases the likelihood that they can find the relational pattern and apply it to a new configuration. The dotted lines in **Figure 1B** denote the one-to-one alignment between elements in the two examples; through these alignments, children may start to "see", for example, that although the biggest square of the left configuration is perceptually different from the biggest circle in the right configuration, they both stand in the same relation to the other members of the array. Perceptual properties or added components that disrupt the alignment of elements disrupt the discovery and generalization of the relational pattern (Kaminski and Sloutsky, 2013; McNeil et al., 2009; Paik and Mix, 2008; Uttal et al., 2008; Rattermann and Gentner, 1998; Son et al., 2012a; Yuan et al., 2017). For example, as shown in **Figure 1C**, the medium square of the left configuration can be mapped either to the largest square of the right configuration (because they are perceptually identical: Object match) or to the medium square of the right configuration (because they are both the medium one within each triplet: Relational match). Likewise, finding the common relational structure is also more difficult when the component elements are heavily detailed and perceptually rich. As shown in **Figure 1D**, the perceptual richness of the Mickey

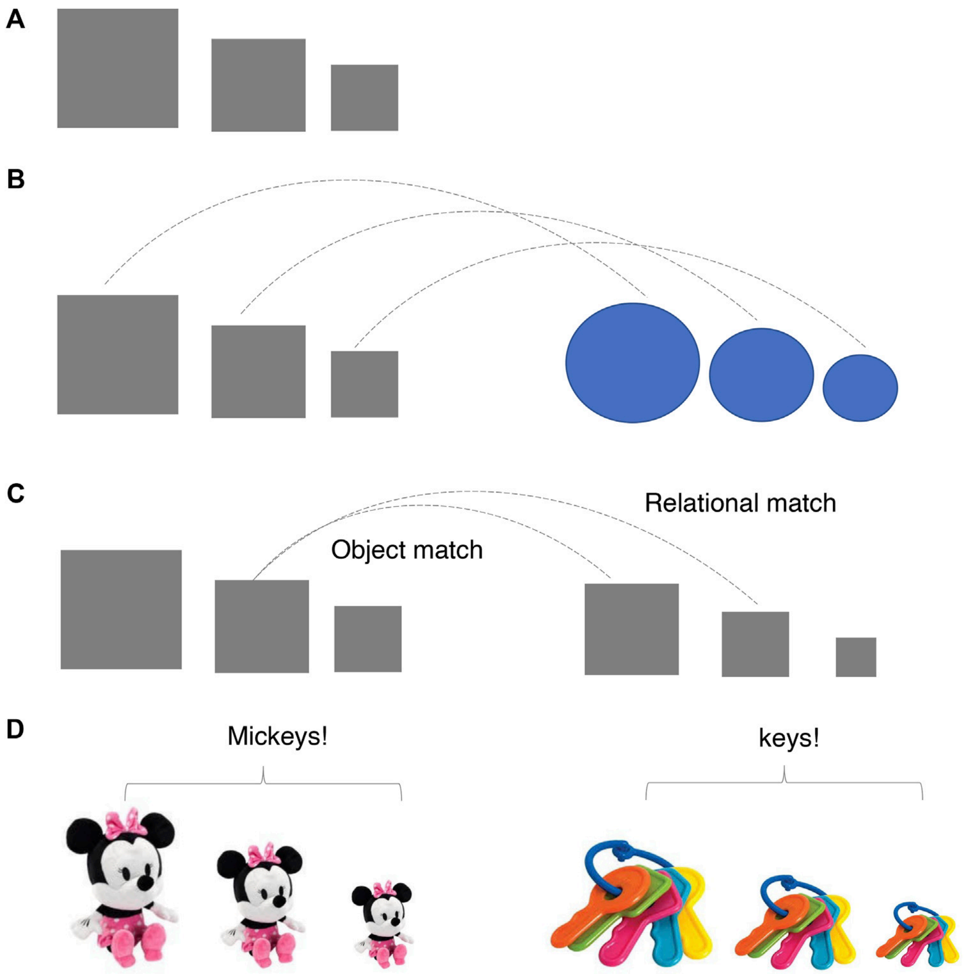


FIGURE 1 | The challenges and ways to extract relational structures: **(A)** Extracting relational patterns from a single exemplar is difficult for young children. **(B)** Aligning and comparing two analogs that share the same relational pattern can help. **(C)** Alignment based on relational match is difficult to establish with the presence of object match. **(D)** Richly detailed components often draw learners' attention to object identities rather than relational patterns.

mouses and the keys with all of their vivid color and interesting shapes can make the individual categories so salient that children represent the left configuration as "They are Mickeys!" rather than "These Mickeys are ordered from largest to smallest". Consistent with this view, past research (McNeil et al., 2009) showed that teaching children calculation with mathematical manipulatives that resembled real money with many details was significantly less effective than using white paper money with only written numbers on them without other extraneous perceptual features.

From the perspective of the Structure Mapping Theory and as illustrated in **Figure 2**, base-10 blocks present a web of mappings at multiple levels all at once, an approach that is contrary to the existing evidence that too many mappings may obscure the discovery of relational structure. The demand of multiple levels of mappings in visual illustrations—meant to clarify text—has been shown to be unhelpful even for college students' learning (Wills et al., 2008; Okeefe et al., 2014; Rau, 2017). This problem of "coordinating multiple representations" is well recognized in the learning science literature on teaching and learning in higher education (Rau, 2017), but it is rarely discussed for children's early learning. This problem is also compounded—for example, in the case of base-10 blocks—when there are features that draw learners' attention to the properties of individual elements at the expense of highlighting the relations among the elements. For example, marking the individual 100 small cube units in a large block (or the "flats") focuses too much on faithfully grounding the "hundred" unit to 100 discrete entities rather than highlighting the relation that the "hundred" place is larger than the "decade" place (which is further larger than the "unit" place). Past research has shown that perceptual properties that contain too many details about individuals often result in learners' failures in recognizing and learning about the relations among individuals and that subtle changes in the direction of presenting the "skeleton" of the relational structure can benefit learning (Rattermann and Gentner, 1998; Paik and Mix, 2008; Kaminski and Sloutsky, 2013; Yuan et al., 2017).

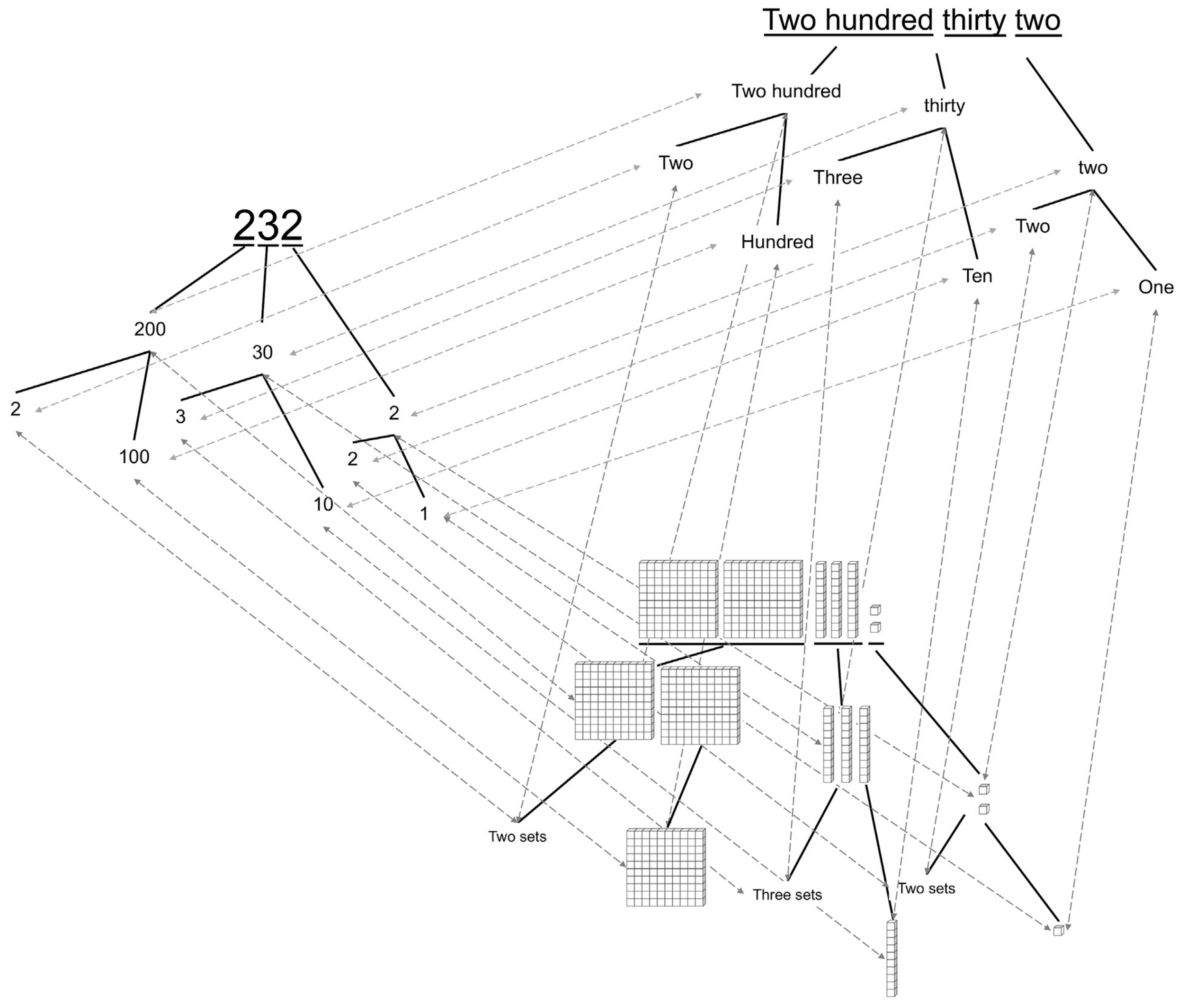


FIGURE 2 | Systems of relational mappings among a written multi-digit number, its corresponding number name, and the base-10 blocks representation of its numerical quantity. The solid lines highlight the structure within each representation, while the dotted lines denote the myriad relational mappings between corresponding components.

From the perspective of the Structure Mapping Theory, relations are found through aligning arrays with the same relational structure and do not necessarily require that the aligned arrays include physical models. Thus, the alignment of number names (one symbolic form) to written forms (another symbolic form)—with no physical model—could be sufficient for an early learner (Mix et al., 2019; Yuan et al., 2020). Multi-digit number names and the written forms have distinct surface forms, but the same underlying relational structure as shown in **Figure 2**. Thus, mapping multi-digit number names to their written forms could yield the discovery of the relational structure—that there are places ordered by relative magnitudes from left to right. This could work because multi-digit number names likely have intuitive meanings that young children partially know—e.g., that "hundred" means "a whole lot" and "-ty" signals a pretty big number as well. These intuitive meanings do not have to be exact to help children find the relational structures. For example, given the big-medium-small relational pattern in **Figure 1A**, children are helped in finding and generalizing that pattern when the elements are aligned with the words "daddy-mommy-baby" (Kotovsky and Gentner, 1996), which only roughly imply size. These arguments, however, do not mean that physical analogies cannot help, they should if they help children align the elements in spoken and written number names.

RATIONALE FOR THE THREE EXPERIMENTS

The three experiments examine the role of relational mapping and physical models in children's discovery of place value through the alignment of multi-digit number names, written notation, or physical models. Across all three experiments, there are pretests and posttests, as well as mapping experiences in between those tests. The mapping experiences merely coach the alignment of elements across examples (spoken number names, written numbers, and physical models) with the same relational structures. Because alignment has been shown to support the extraction of the relational structure in a few trials (Son et al., 2012a; Son et al., 2012b), there were not many mapping trials. In addition, explanations are minimal—typical in relational learning experiences—since aligning elements is the hypothesized key factor to discovering relations. The participants were 4- to 6-year-olds, who were in preschool or kindergarten classes and who had not been formally introduced to the place value system in school.

Experiment 1

Experiment 1 was designed to test whether the discovery of the common relational structure of multi-digit number names, their written forms, and their magnitudes is better achieved—through the use of physical models (base-10 blocks and abacus) which provide groundings of the base-10 hierarchies (e.g., 100 is 10 sets of 10), or merely by mapping the two symbolic forms (written and spoken) of multi-digit numbers. In the experiment, the *Symbols-to-Symbols* condition involves only number names and their corresponding digits with the mapping goal consisting of a direct alignment between corresponding elements of the two symbol systems. Next in complexity is the *Symbols-to-Abacus* condition; in addition to number names and written forms, the columns of the abacus align spatially with the places and the numbers of discs at each column align with the digit in each place, providing the common meaning that links the written symbols and their spoken names. The *Symbols-to-Blocks* condition is the most complex; the symbols align with the number of whole blocks, the cells within the blocks align with the actual amount represented, and the spatial arrangement of the blocks aligns with different places. If the number of correlated features supports finding the aligned relation, one might expect the *Symbols-to-Blocks* condition to best support the relational structure underlying number names, their written forms, and their relative magnitudes. If simplicity and alignment of common elements is the key, then the *Symbols-to-Symbols* condition may lead to better discovery of places and their relative magnitudes.

In all conditions, we used a coached imitation task to foster alignment of heard number names, written digits, and physical models (in the physical model conditions). The experimenter said the number name, created the written number with digit cards, and then in the model conditions, made a model of the number with the manipulatives, repeating the number name, and aligning the model and the written number in space. The child was then asked to copy these constructions with the experimenter's model and digit cards in view, and if the child made a mistake, he or she was coached to make the correct constructions by the experimenter. We measured and used the number of errors in the mapping task as an indicator of the perceptual transparency of the alignments to the children. We also measured children's pre- and post-test performance using numbers that were not trained during mapping. There were two pre- and posttest tasks: mapping number names to written digits and magnitude comparison of written numbers.

Participants

Seventy-five children (37 females and 38 males, age range: 4.03–6.88 years) participated in the study. As noted above, the participating children in this and all following studies were enrolled in preschool or kindergarten; none had entered first grade. The age of the sample was sensibly and positively skewed (median: 5.33 years, mean: 5.44 years), reflecting a range of 4 to early 6 years, with very few older 6-year-olds who had not yet started first grade at the time of the study due to various reasons (e.g., their birthday fell after the school district's fall cut-off for 6-year-olds to enter first grade). This age range was appropriate, given that we were broadly interested in children's early learning of multi-digit numbers before formal education. Participants were recruited through community organizations (e.g., farmers' markets, child outreach events, boys' and girls' clubs) and local preschool and daycare centers. The sample of children was broadly representative of the local population (84% European American, 5% African American, 5% Asian American, 2% Latino, 4% Others) and consisted of predominantly working- and middle-class families. The study was approved by the Human Subjects and Institutional Review Boards at Indiana University. In this and all following studies, informed consents were obtained from the legal guardian and assents were obtained from the children prior to the experiment. Children were randomly assigned to one of three conditions: Symbols-to-Symbols mapping (n = 27), Symbols-to-Abacus mapping (n = 23), and Symbols-to-Blocks mapping (n = 25).

Materials and Procedure

This experiment had three phases: pretest, relational mapping trials, and posttest.

Pre- and Posttest

The pre- and posttests consisted of two established tasks: the which-N and which-More tasks (Mix et al., 2014; Yuan et al., 2019). On each trial, children were presented with a pair of written multi-digit numbers. In the Which-N task, children were told a spoken number name and then asked to select the written form that matched the name; in the which-More task, they were asked to select the one that was more. There were 16 trials for each of the tasks with a total of 32 trials (**Table 1**). Accordingly, 32 cards were made with two multi-digit numbers (roughly 17.78 cm wide and 12.7 cm tall) printed at the center of the card. The particular numbers used in the tasks were randomly sampled from 1- to 4-digit numbers. The pair of target and foil numbers were chosen from a variety of different types to avoid the possibility that knowing any single strategy or heuristic would allow the participant to solve all (or majority) of the trials. For example, simply knowing that numbers with more digits signify larger values is not enough to successfully choose the larger value between 223 v 220. Similarly, knowing that "three hundred and five" should start with "3" alone is not enough to choose the correct written form of "three hundred and five" given 350 v 305. These different types have been used in previous research (Yuan et al., 2019) and include single digits numbers (e.g., 2 v 8), numbers with different numbers

TABLE 1 | All tasks and items used in the three experiments.

	Pre- and Post-tests	Items
Exp 1	The Which-N task (16 trials)	2 v 8, 11 v 24, 12 v 22, 15 v 5, 36 v 306, 64 v 604, 85 v 850, 105 v 125, 201 v 21, 206 v 260, 350 v 305, 402 v 42, 670 v 67, 807 v 78, 1000 v 100, 1002 v 1020
	The Which-More task (16 trials)	3 v 7, 6 v 8, 11 v 19, 14 v 41, 16 v 62, 26 v 73, 30 v 60, 72 v 27, 82 v 081[a], 100 v 10, 101 v 99,123 v 321, 223 v 220, 585 v 525, 670 v 270, 4620 v 4520
Exp 2	The Which-N task (20 trials)	Easy items (10 trials): 2 v 8, 12 v 22, 14 v 41, 15 v 5, 24 v 11, 64 v 604, 67 v 670, 125 v 105, 350 v 305, 900 v 99 Hard items (10 trials): 189 v 198, 362 v 326, 485 v 4085, 677 v 766, 1900 v 1009, 2060 v 2006, 3070 v 307, 5109 v 5910, 7014 v 7804, 8503 v 8350
	The Which-More task (20 trials)	Easy items (10 trials): 3 v 7, 16 v 62, 26 v 73, 30 v 60, 72 v 27, 100 v 10, 123 v 321, 223 v 220, 585 v 525, 670 v 270 Hard items (10 trials): 536 v 5362, 690 v 609, 751 v 571, 899 v 988, 1010 v 101, 2395 v 2315, 4208 v 4820, 6040 v 6400, 5035 v 5605, 7300 v 7003
Exp 3	The Which-N task (16 trials)	2 v 8, 11 v 24, 12 v 22, 15 v 5, 36 v 306, 64 v 604, 85 v 850, 101 v 100, 102 v 120, 105 v 125, 201 v 21, 206 v 260, 350 v 305, 402 v 42, 670 v 67, 807 v 78
	The Make-a-model task (6 trials)	5, 8, 50, 73, 429, 601
	The Choose-a-model task (16 trials)	3 v 7, 6 v 8, 11 v 19, 14 v 41, 16 v 62, 26 v 73, 30 v 60, 72 v 27, 82 v 81, 100 v 10, 101 v 99, 123 v 321, 223 v 220, 462 v 452, 585 v 525, 670 v 270

This item was originally added as a catch on for the strategy of just counting the number of digits. Many children indeed utilized this strategy. Since this trial contained a non-existing number, it was not included in the reported analysis. Additional analysis including this trial did not change the pattern of results.

of places (e.g., 36 v 306), transpositions (e.g., 350 v 305), numbers that differed in only one digit (e.g., 105 v 125), and numbers with no digit overlapping (e.g., 11 v 24). Items in the which-More task were sampled using the same method as those used in the which-N task, but the two tasks involved different numbers to avoid the possibility that exposure to items in one task would influence participants' responses to the same items in the other task. Because during the which-More task children were expected to always choose the numerically larger number, consistent with established procedures and to counterbalance the response demand across the two tasks, the experimenter always asked the numerically smaller number of the two in the which-N task.

For each of the test tasks, the same items were used at both pretest and posttest, which allowed us to measure the effectiveness of relational mapping on children's discovery of the relational structure. Two versions were created that counterbalanced the order of the individual items. If a child received version A at pretest, she/he would receive version B at posttest; we did this to minimize the test-retest effect. No feedback was given at pretest or posttest.

Mapping

As shown in **Figure 3** top, during the mapping trials, participants imitated the experimenter to make multi-digit numbers using three different sets of materials (i.e., number cards only, number cards with abacus, number cards with base-10 blocks). The base-10 blocks included three types that each represents one place value unit. The small squares were 1.016 cm wide and 1.016 cm tall and were used to represent the unit of one. The bars were composed of ten small squares attached together to form a bar shape (1.016 cm wide and 10.16 cm long), and they represented the unit of ten. The big squares were composed of ten bars attached together to form a big square (10.16 cm by 10.16 cm), and they represented the unit of hundred. The abacus was 28 cm wide and 25 cm tall, composed of two horizontal bars at the top and bottom and three vertical bars for holding the discs at each place value unit (e.g., left bar: the hundred place, middle bar: the decade place, right bar: the unit place). Nine discs were attached to each of the vertical bars and different colors were used to help participants differentiate the places: discs in the hundred place were blue, those in the decade place were yellow, and those in the unit place were red. All of the discs were of the same size (with a diameter of 3 cm). The number cards (4 cm wide and 7 ¼ cm tall) were composed of individual cards with single digits (i.e., 0–9) printed at the center of each card in Times font. There were 15 mapping trials during which children from all three conditions were first presented with a target card (12 cm wide and 7 ¼ cm tall) with the target multi-digit number printed at the center in Times font. The 15 target numbers were: 14, 163, 187, 65, 4, 23, 451, 52, 6, 673, 72, 8, 901, 838, 94. The multi-digit numbers that participants received during mapping and those used for pre- and post-tests were different. In other words, if participants' performances in the testing tasks have improved from pre- to post-test, then it would suggest that they have learned from the mapping experience and generalized that learning to different items used in the testing tasks.

For all three conditions, 10 piles of individual digits cards from "0" to "9" were laid out in sequence on the left side of a table. During the mapping trials, both the experimenter and the child took individual cards from these piles to make multi-digit numbers. In each condition, there were 15 mapping trials; this small number is consistent with prior work which has shown that a relatively few such mapping experiences can yield generalizable discovery of relations in preschoolers (Rattermann and Gentner, 1998; Son et al., 2012a; Yuan et al., 2017).

Symbols-to-Symbols mapping. The experimenter first introduced the individual number cards to the participant, "We have some number cards here." She then pointed to each pile from "0" to "9" and named the digit for the child from "zero" to "nine". She told the child that they were going to make some numbers using these cards. She gave one example by picking up a card with "1" and a card with "3", laying them down on the table side-by-side while saying, "I have made thirteen. Can you make thirteen using these cards?" There were 15 such mapping trials. Children only had to copy the same actions as the experimenter, so potentially performance could be

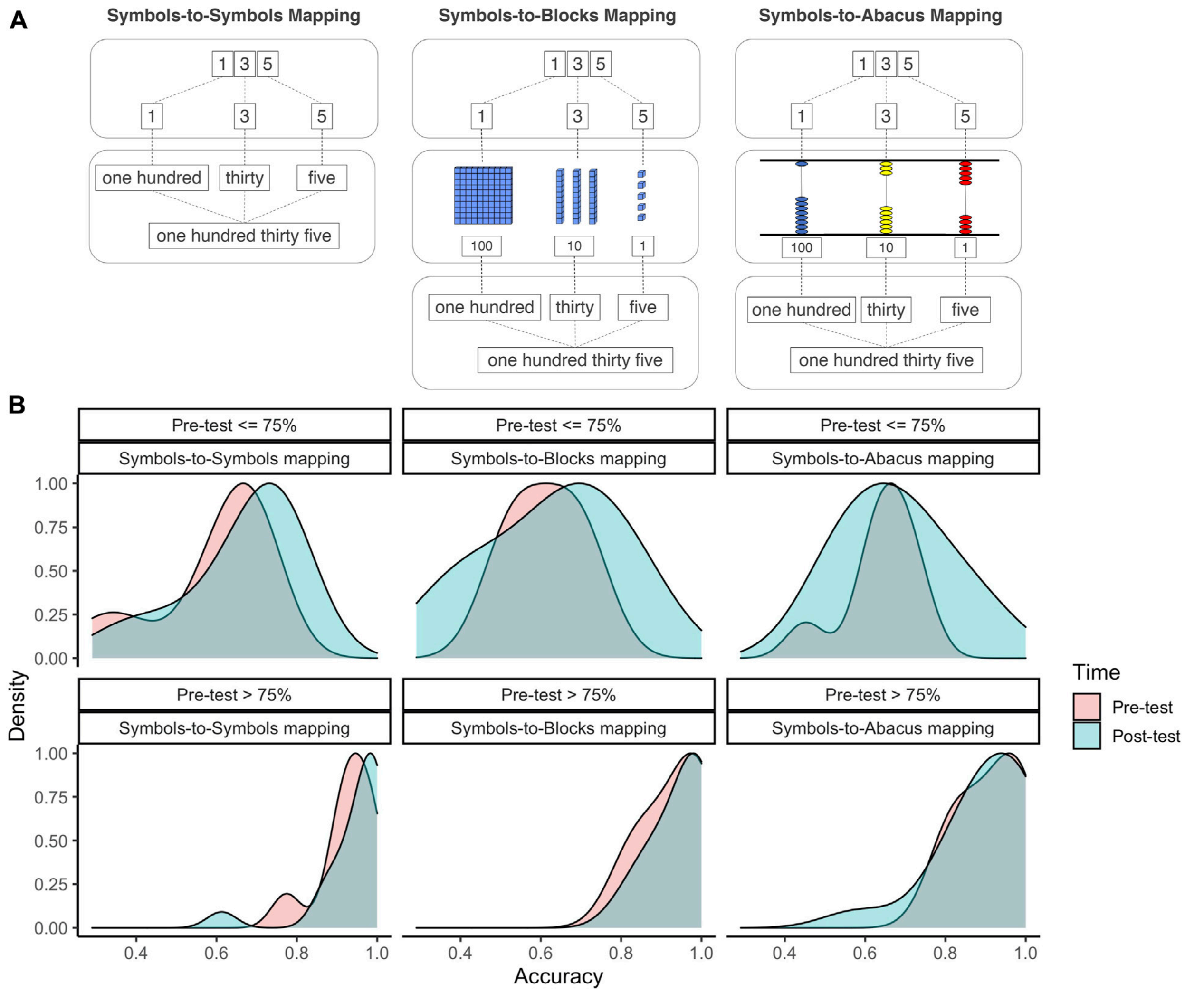


FIGURE 3 | (A): Illustrations for the three mapping conditions in Experiment 1. **(B)**: Density graphs of participants' accuracy in Experiment 1 by condition, test time (pre- and post-test), and pre-test familiarity with multi-digit number symbols. In **(B) top row:** children who scored less than or equal to 75% at pre-test. In **(B) bottom row:** children who scored above 75% at pre-test. Pink indicates pre-test performance and turquoise indicates post-test performance.

errorless. The ability to make correct copies—trial by trial—provides a measure of the obviousness of the alignment between different components (e.g., number names to written forms). In all conditions, initial copy attempts by the child were scored as correct or incorrect (e.g., if the to be copied item was "13" and the child took a 7 instead of a 1, or made 31 instead of 13, it would be scored as incorrect). On all incorrect trials, the experimenter coached the child into making a final correct copy and repeated the name (e.g., "See you made 13") to the correct version of the written form.

Symbols-to-Abacus mapping. The experimenter introduced and named the digits cards from "0" to "9" similar to the Symbols-to-Symbols condition. She then familiarized the participant to the abacus, telling the children the name of the abacus and showing them how the discs could be moved and allowing them to do so. The experimenter then gave an example by laying down the cards "1" and "3" while saying "Here is thirteen. Now watch, I am going to make thirteen using the abacus." She then put the correct abacus configuration down underneath the cards while saying, "I'm making thirteen," or "Here is thirteen." She then asked the child, "Can you make thirteen using your cards and then with the abacus?" The participant then made the corresponding number first with their own cards and then with the abacus and the experimenter repeated the spoken name when the written number was correctly formed and when the abacus model was correctly formed. On all incorrect trials, the experimenter coached the child into making the final correct copy and repeated the correct number name.

Symbols-to-Blocks mapping. The experimenter first introduced and named the digits cards from "0" to "9" in the same way as the Symbols-to-Symbols condition and introduced the different sized blocks. During the mapping trials, the experimenter first picked out individual cards and laid them down on the table to make a target number. For example, she might pick out cards "1" and "3", laying them down side-by-side on the table while telling the participant that "I have made thirteen using our cards. Next, I am going to make thirteen using our blocks." She then put the correct number of blocks down underneath the cards (a tens block under the card "1" and three ones blocks under the card "3") while saying, "I'm making thirteen," or "Here is thirteen." She then asked the child, "Can you make thirteen using your cards and blocks?" The participant then made the corresponding number first with their own cards and then with their own blocks. The experimenter repeated the number name when the written form was correctly formed and also when the block model was correctly formed. On all incorrect trials, the experimenter coached the child into making the final correct copy and repeated the correct number name.

Results

To determine the possibility of pre-existing group differences, we compared children's performance at pre-test across the three conditions. In this and all following analyses, logistic mixed effect models were used to evaluate children's performance under different training conditions. Such models allowed us to utilize item-level data from each participant (rather than computing a summary score—e.g., mean, median—for each participant) and to take into consideration the multi-level hierarchical structure of experimental studies, in which multiple trials are nested within corresponding participants and multiple participants are nested within corresponding experimental conditions (Singmann and Kellen, 2019). A logistic mixed effect model was conducted in the R environment (R Core Team, 2020) using the Afex package (Singmann et al., 2015). Condition was entered as a fixed variable, participant and test item were entered as random variables; the dependent variable was the accuracy of the individual trials (i.e., 0 or 1) in the which-N and which-More tasks at pre-test. There was no significant main effect of condition, χ^2 (2) = 0.71, p = 0.70, suggesting that children assigned to the three mapping conditions started out at comparable levels of competencies. As evident in **Figure 3**, there were, however, considerable individual differences with some children performing at levels above 75% at pretest but other children performing much more poorly. These individual differences in early informal knowledge have been reported previously (Byrge et al., 2014; Yuan et al., 2019). Although continuous age (in month) was correlated with individual children's performance at pretest (r^2 = 0.42, p < 0.001) and post-test (r^2 = 0.38, p < 0.001), there was no evidence that age was related to how much children had improved after the training (overall: r^2 = 0.0005, p = 0.85; Symbols-to-Symbols condition: r^2 = 0.04, p = 0.29; Symbols-to-Blocks condition: r^2 = 0.02, p = 0.54; Symbols-to-Abacus condition: r^2 = 0.007, p = 0.69). As stated above, none of the participants had yet attended first grade at the time of the study. Given that we were broadly interested in how children can learn from symbolic and physical representations of multi-digit numbers before receiving formal place value instructions at schools and the large individual differences in early knowledge beyond the factor of age, we collapsed children across the different age groups in all following analyses.

Children's ability to copy the experimenter's models (and the need of direct coaching) during the mapping trials provides a measure of the obviousness of the alignment of elements across arrays. Children were reasonably successful in the copying task across the three conditions but were better able to correctly copy the arrays in the Symbols-to-Symbols conditions than in the two physical models conditions. A logistic mixed effect model, in which condition was entered as the fixed variable, participant and test item were entered as random variables, revealed a significant main effect of condition, χ^2 (2) = 33.46, p < 0.001. Children were significantly more accurate during mapping trials in the Symbols-to-Symbols condition (Mean = 0.98, SE = 0.008) than in the Symbols-to-Blocks condition (Mean = 0.78, SE = 0.039), t (26) = 4.77, p < 0.001, and the Symbols-to-Abacus condition (Mean = 0.87, SE = 0.036), t (24) = 3.16, p = 0.004. There was no significant difference between the Symbols-to-Blocks and Symbols-to-Abacus conditions, t (46) = 1.38, p = 0.17.

The key test of the discovery of the relational pattern is whether children can apply the pattern to arrays that were not experienced during training (measured in the Which-N task) and whether they can make inferences from the relational patterns as to the indicated magnitude (measured in the Which-More task). Children's performances relative to pretest increased in the Symbols-to-Symbols condition but not in the Symbols-to-Blocks or the Symbols-to-Abacus conditions. Logistic mixed effect models on the pre- and post-test performances were conducted for each condition with test time (pre-test vs. post-test) and test task (which-N vs. which-More) entered as fixed variables and participant and test item entered as random variables, and accuracy on individual items as the dependent variable. For the Symbols-to-Symbols condition, there was a significant main effect of test time, χ^2 (1) = 4.24, p = 0.039, with performance improving from pre-test (Mean = 0.80, SE = 0.04) to post-test (Mean = 0.84, SE = 0.03). There was no reliable main effect of test task, χ^2 (1) = 1.72, p = 0.19, nor interaction between the two fixed effects, χ^2 (1) = 0.13, p = 0.72. For both the Symbols-to-Blocks condition and the Symbols-to-Abacus condition, the models failed to detect any significant main effect of test time (ps > 0.43), task (ps > 0.14), nor an interaction between them (ps > 0.71).

Figure 3 bottom shows density plots of children's performance at pre-test and post-test for the three conditions and separated by children with high and low prior knowledge (defined by 75%[1] accuracy in the composite score of the which-N and which-More task at pre-test). In the Symbols-to-Symbols condition, the performance distributions for both children with high and low prior knowledge have shifted to the right from pre- to post-test. Interestingly, for both the Symbols-to-Abacus condition and the Symbols-to-Blocks condition, the distribution of children with low prior knowledge widened after the mapping experience, suggesting that the use of physical models helped some children but hurt others. Past work has shown that relational structures become easier to perceive with expertise and exposure to the content domain (Chi, 1978; Chi et al., 1981). Thus, one possible explanation for the ineffectiveness of traditional manipulatives in the current experiment is that most children were not ready for and could not yet utilize the information in these more complex (albeit more accurate) models of the notational system. We return to these issues in Experiment 3 and in the General Discussion.

[1]Using median-split as a grouping method does not change the pattern of results

Discussion

Experiment 1 shows that 1) the alignment of number names and written forms is sufficient for children to discover the relational patterns that map number names to written forms and to judge the relative magnitudes of multidigit numbers, and 2) it illustrates how adding additional information—even though relevant and even though redundant—can make the ***initial*** discovery of the relational patterns through structure mapping less likely.

Experiment 2

The finding that a few Symbols-to-Symbols mapping experiences supported the discovery and generalization of the relational pattern underlying number names and their written forms is surprising in the context of a body of literature that has generally concluded that the mapping between number names and written forms is hard to learn (Baroody, 1990; Fuson, 1990). Accordingly, the explicit goal of Experiment 2 was to replicate with a larger sample the finding that relatively few mapping trials enabled learners to find and generalize the pattern. To this end, we realized the mapping trials in two different ways: using the approach of Experiment 1 and also a slightly different approach that has been commonly used in Montessori schools. In the Digits mapping condition as in Experiment 1, children created multi-digit numbers by using individual digit cards (e.g., “2”, “3”, “2”) mapped to their spoken name “two hundred”, “thirty”, “two”). In the Expanded mapping condition, they mapped expanded cards (e.g., “200”, “30”, “2”—“two hundred”, “thirty”, “two”) to the name by stacking them on top of each other to form a visual array showing the place value notation (“232”).

Participants

Ninety-three children (50 females) recruited from the same general population as Experiment 1 participated in this study (age range 4.03–6.82 years, median: 5.26, mean: 5.39). Children were randomly assigned to the Digits mapping condition ($n = 42$) or the Expanded mapping condition ($n = 51$). The study was approved by the Human Subjects and Institutional Review Boards at Indiana University. Informed consents were obtained from the guardian and assents were obtained from the children prior to the study.

Materials and Procedures

The which-N and which-More tasks were used as Pre- and Post-tests like Experiment 1. There were 20 trials for each of the tasks (shown in **Table 1**) with 10 easier trials that involved mostly 2- and 3-digits numbers and 10 harder trials that involved 3- and 4-digit numbers. Children were first given the 10 easy trials; if they got 7 out of 10 of those trials correct, we proceeded onto the next set of 10 trials. We used this approach to maintain the participation of children who found the tasks (particularly at pre-test) too difficult to continue through all 20 trials. Two orders were created that counterbalanced the order of the individual items across pre- and post-tests. No feedback was given at pre- and post-tests.

Mapping

There were 20 mapping trials and correspondingly 20 cards with the numbers 163, 51, 846, 47, 271, 94, 18, 36, 328, 451, 62, 65, 719, 653, 594, 587, 23, 89, 72, 972. The mapping experiences immediately followed the pre-test for all participants. In both conditions, the procedures were structured similarly as in Experiment 1 except for two differences. First, we followed a “progressive alignment” approach (Thompson and Opfer, 2010; Gentner et al., 2011) often used in relational mapping studies that began with 10 mapping trials involving two-digit numbers and followed by 10 mapping trials with three-digit numbers. Second, because the Expanded mapping condition involved all possible numbers from single to three-digit numbers (i.e., 1–900), laying out all of these components (as in Experiment 1) in front of the child was not feasible. Instead, on each mapping trial, the experimenter had a set of individual cards needed for assembling the target number. She also gave the child an identical set of cards. For example, as shown in **Figure 4**, if the target mapping number was 135 and the child was in the Expanded mapping condition, both the experimenter and child would have cards 135, 100, 30, and 5. The 3-digit number cards were 12 cm wide and 7 ¼ cm tall; the 2-digit number cards were 8 cm wide and 7 ¼ cm tall; the 1-digt number cards were 4 cm wide and 7 ¼ cm tall. If the child was in the Digits mapping condition, they would both have the cards 135, 1, 3, and 5. Since all of the cards were single digit in this condition, they all have the dimension of 4 cm wide and 7 ¼ cm tall. The experimenter always handed the child his or her set of cards at the beginning of each trial. She then demonstrated how to make the number, had the child copy her action after each step, and scaffolded if needed.

Results

Pre-test performance in the Digits and Expanded mapping conditions did not differ as indicated by a logistic mixed effect model—in which condition was entered as a fixed variable, participant and test item were entered as random variables, and the dependent variable was the accuracy of the individual trials (i.e., 0 or 1), χ^2 (1) = 0.12, $p = 0.73$. Again, as shown in **Figure 4**, there were considerable individual differences in pre-test performance as some children performed very well and others quite poorly at pre-test. During the mapping task when children were asked to imitate the experimenter in making the written forms in response to hearing the name, children readily imitated the experimenter with very few errors (overall 99% correct in both conditions), thus readily discovering the relations between number names and written digits. Children in both conditions also performed better at post-test than pre-test on untrained digits indicating generalization of the learning. To ask whether the advantage of one mapping condition was higher than the other, we entered both conditions in one logistic mixed effect model. Condition and test time were entered as fixed variables, participant and test item were entered as random variables. The model detected a significant main effect of test time, performance improved from pre-test to post-test, χ^2 (1) = 11.27, $p < 0.001$, while there was no significant main effect of condition, χ^2 (1) = 0.06, $p = 0.81$, nor condition and test time interaction, χ^2 (1) = 0.08, $p = 0.77$. Both of the Expanded mapping condition (pre-test Mean = 0.76, SE = 0.02, post-test Mean = 0.79, SE = 0.02) and the Digits mapping condition (pre-test Mean = 0.75, SE = 0.02, post-test Mean = 0.78, SE = 0.03) improved from pre- to post-test. As shown in **Figure 4**, children improved from pre- (Mean = 0.75, SE = 0.005) to post-test (Mean = 0.78, SE = 0.004) after the mapping experience, regardless of their conditions.

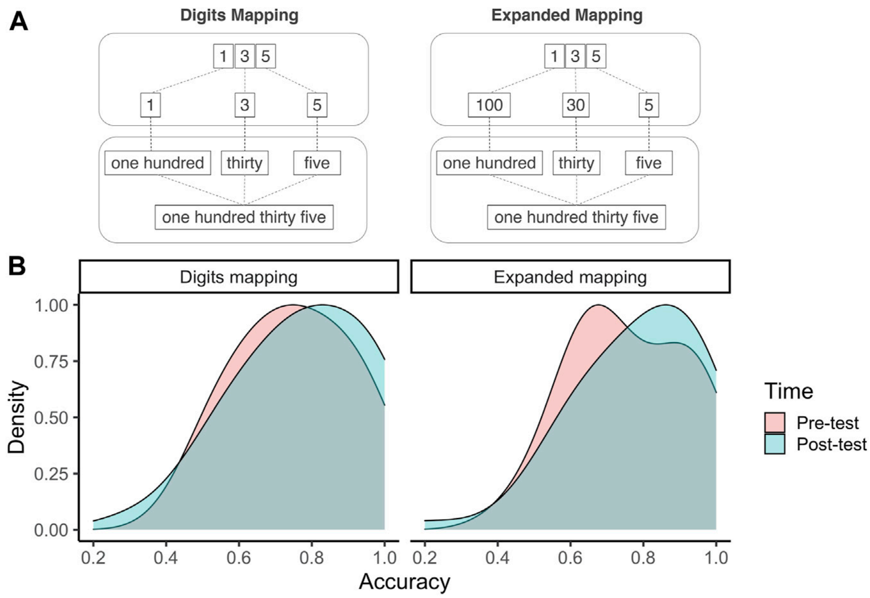


FIGURE 4 | (A): Illustrations for the two mapping conditions in Experiment 2. **(B)**: Density graphs of participants' accuracy in Experiment 2 by condition and test time (Pre- and Post-test). Pink indicates pre-test performance and turquoise indicates post-test performance.

Because of the large individual differences and strong performance of some children at pre-test, we also repeated the analyses using only the data from children ($N = 50$) who were correct less than or equal to 75% of the trials on the combined which-N and which-More tasks at pre-test. Condition and test time were entered as fixed variables, participant and test item were entered as random variables. The model detected a significant main effect of test time, performance improved from pre-test to post-test, χ^2 (1) = 5.69, $p = 0.017$, while there was no significant main effect of condition, χ^2 (1) = 0.39, $p = 0.53$, nor condition and test time interaction, χ^2 (1) = 0.01, $p = 0.92$.

Discussion

The post-test required children both to generalize the relational structure to new number names and written forms and generalize their learning about the relational mappings to judgements of magnitude—a task on which they were given no experience during the relational mapping experiences and so children had to use the discovered relational structure in a new way. The effects of the brief mapping experience were small and were not proposed as a sufficient training procedure in and of themselves. However, they indicated the potential value of reconceptualizing the *initial* learning problem as one of discovering relational patterns, and the results showed that relational mapping from Symbols (heard names)-to-Symbols (written forms) is useful in this domain just as in other cognitive domains (McNeil et al., 2009; Kaminski and Sloutsky, 2013).

Experiment 3

Although the results from Experiment 1 did not seem to support the idea that early knowledge about multi-digit number symbols can be learned via grounding to physical objects, it is important to further determine why and how before making a sweeping generalization. If analogical mapping is the core mechanism that drives the positive results of the Symbols-to-Symbols mapping in Experiment 1 and 2, then mapping symbols to physical models should be helpful *if* they direct the learners' attention to the relevant relations (Gentner and Toupin, 1986; Goldstone, 1998; Jee et al., 2010; Jee et al., 2013). The two physical model conditions in Experiment 1 may have failed to do so. For instance, as shown in **Figure 2**, the number of mappings between base-10 blocks and number symbols may be too much, too distracting, and not focused on the critical early knowledge—e.g., there are different places representing different relative magnitudes—that children need to know when first learning the multi-digit system.

To test the above hypothesis, in Experiment 3, we chose to focus on the abacus as a physical model. We chose the abacus because the relational structure of the abacus is analogy-like in that it does not represent the exact magnitudes (as in Base-10 Blocks) but instead represents the system as the counts of units (the discs) in the different places. To directly test the role of physical models and to reduce the complexity of the alignments (a potential problem for the physical models conditions in Experiment 1), the mapping experiences were from the spoken names to the physical models. The written forms were not used in the mapping experiences but were included in the pre- and post-tests. Success at post-test thus required generalization of experienced mapping (heard name to physical model) to a new mapping between spoken number names to the written forms (the which-N task used in Experiments 1 and 2).

We constructed three "abacus" conditions as shown in **Figure 5**: 1) Standard abacus: the original abacus as in Experiment 1, 2) Sized abacus: the original abacus to which we added a redundant place cue in which the discs varied in

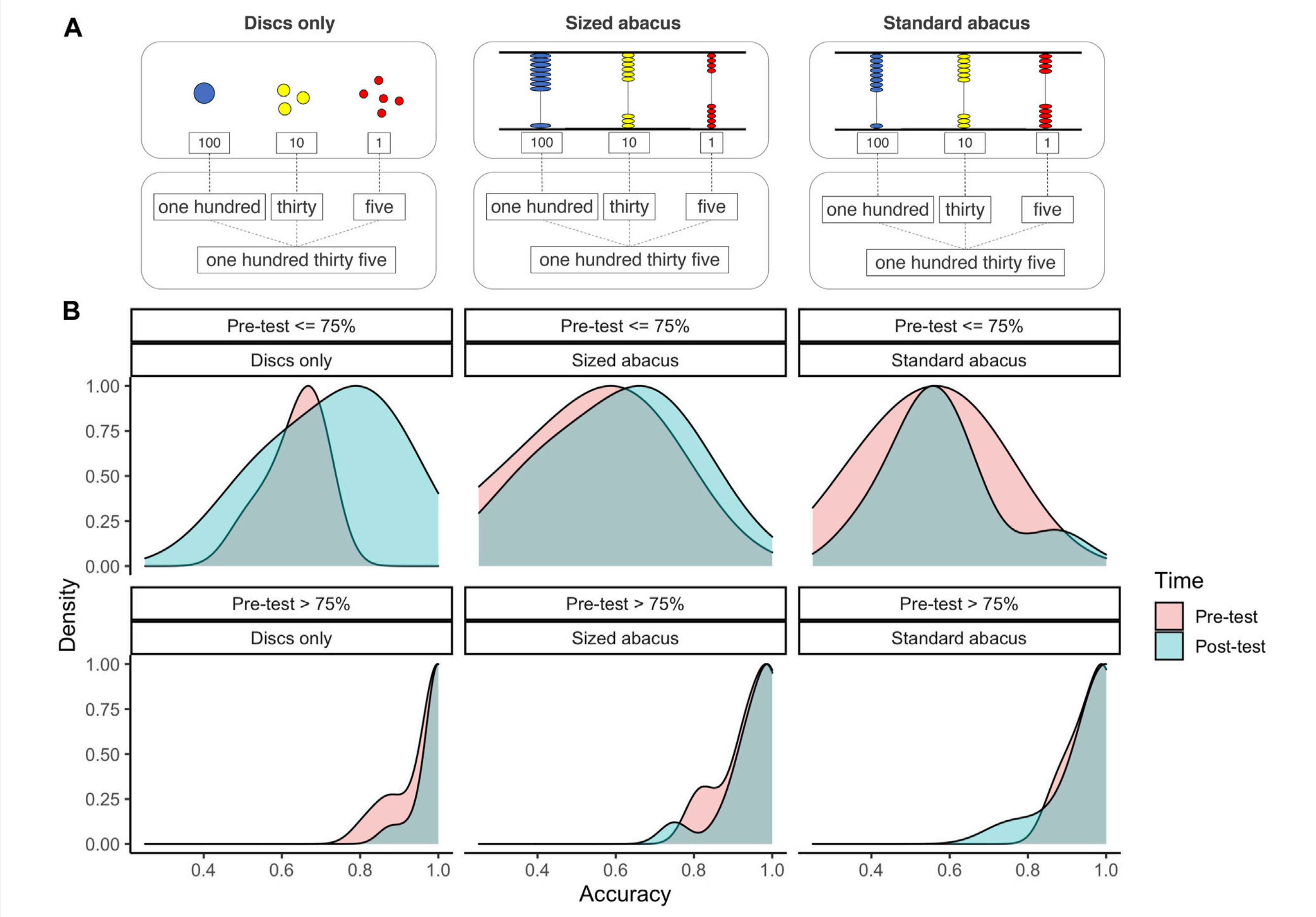


FIGURE 5 | (A): Illustrations for the three mapping conditions in Experiment 3. **(B)**: Density graphs of participants' accuracy in Experiment 3 by condition, test time (pre- and post-test), and pre-test familiarity with multi-digit number symbols. In **(B) top row:** children who scored less than or equal 75% at pre-test. In **(B) bottom row:** children who scored above 75% at pre-test. Pink indicates pre-test performance and turquoise indicates post-test performance.

size so that the relative magnitude (not exact) was modeled by the relative sizes of the counters; and 3) Discs only: a deconstructed "abacus" that used only discs varying in size and set on the table in separate groups.

The Sized abacus condition is just as complex as the Standard abacus condition, but the different sizes may make the relative magnitudes of the places more mappable by highlighting that property. The Discs only condition takes this potentially greater mappability—bigger discs indicate places that stand for bigger quantity—and simplifies the context of the presentation. The Discs only condition is like the Symbols-to-Symbols conditions of Experiments 1 and 2 in that heard number named were mapped to visual patterns; only here the visual patterns are not written multi-digit numbers but discs that vary in size—representing the relative magnitudes of the quantities represented in each place.

The Discs only condition was also motivated by prior findings on children's proportional reasoning Boyer et al. (2008) As shown in **Figure 6**, one study presented 4- to 6- year-old children with an exemplar (showing a particular proportion) together with two choices and asked the children to find the one that showed the same proportion as the exemplar. Children were more successful with continuous length than with discrete length (**Figure 6A**) in which each individual unit was clearly and perceptually defined. Given discrete representations, the children attended to the individual units, prompting strategies such as counting units, rather than attending to the part-and-whole relation that defines the concept of proportion. Likewise, early "approximate" learning about base-10 concept may benefit more from seeing different sized discs that mark only the main idea—relative magnitudes of different places—compared to base-10 blocks that convey not only the relative magnitudes but also the precise place value principles (e.g., 100 is 100 sets of 1s). As illustrated in **Figure 6B**), the presence of the individual units on a base-10 block may prompt children to count, rather than attending to the relative magnitude of the places (hundred > decade > unit) and the mapping between places to their number names. In contrast, as shown in **Figure 6C**), the use of three simple discs allows for more efficient mapping because the only difference among the discs is their relative size—there is no other feature that would allow an alternative mapping.

In sum, the mapping experiences in Experiment 3 were between multi-digit number names (with place value terms) to a physical

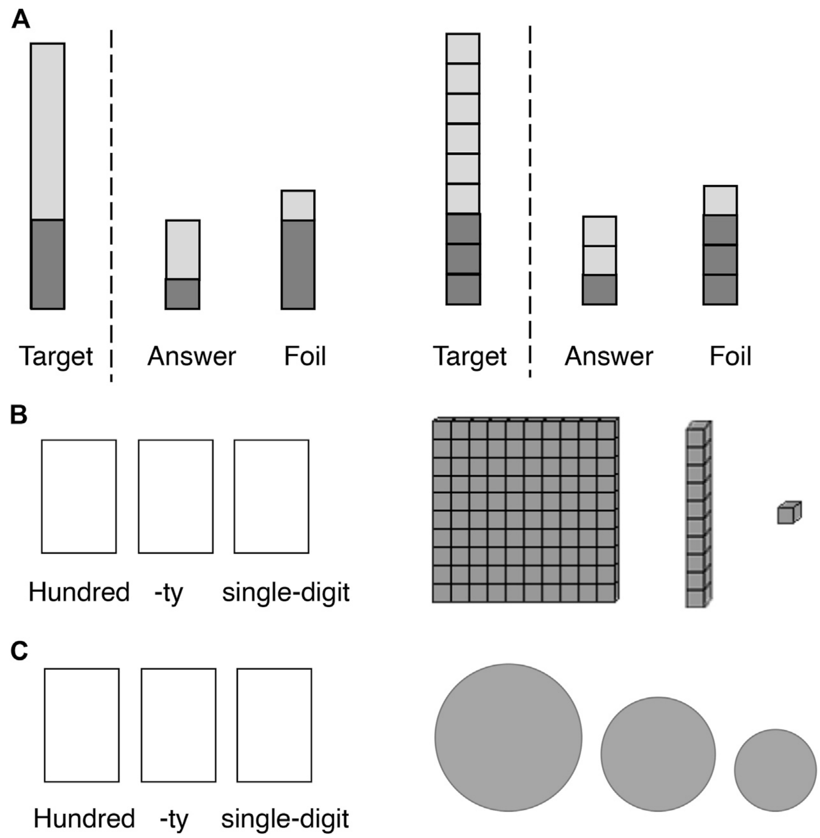


FIGURE 6 | How discrete presentation draws attention to individual units rather than relational patterns and how to better highlight the relational structure of place value. **(A) Left:** the continuous presentation format, **Right:** the discrete presentation format used in a proportional reasoning task (Boyer et al., 2008). **(B)** How base-10 blocks represent the relational structure of place value. **(C)** How simple shapes (e.g., big, medium, and small discs) represent the relational structure of place value.

representation of the numerical magnitude: the Discs only condition used relative *size* (big, medium, small) and the number of discs to represent quantity; the Standard abacus uses *spatial arrangement* (left, middle, right) and the number of discs; the Sized abacus uses both *size* and *spatial arrangement* of discs to represent quantity. The abacus—both standard and sized—also have other details that could be distracting (the top and bottom, the metal poles, etc.). Given the result from Experiment 1, we expect that the Standard abacus condition will not lead to significant learning. The interesting question here is whether the Discs only condition and/or the Sized abacus condition would be more helpful. On the one hand, the redundant cues (size and spatial arrangement) presented in the Sized abacus condition might be beneficial, because it allows for multiple strategies (if one way fails, there is a backup). In contrast, a "less is more" principle to relational discovery suggests that only having one cue—one way to establish the mapping—might be better because it will draw the learner's attention directly to the to-be-discovered relation.

Participants

Fifty-nine children (29 females and 30 males, age range: 4.00–7.04 years, median: 5.39, mean: 5.50) from the same general population as Experiments 1 and 2 participated in this study. They were randomly assigned to three conditions: Standard abacus (n = 19), Sized abacus (n = 19), and Discs only (n = 21). The study was approved by the Human Subjects and Institutional Review Boards at Indiana University. Informed consents were obtained from the legal guardian and assents were obtained from the children prior to the experiment.

Materials and Procedure

The experiment has 4 phases: 1) A pre-test that consisted solely of the which-N task with the choices being written digits (as in Experiments 1 and 2), 2) mapping trials in which children imitated the experimenters in making of the number in one of the three "abacus" conditions, 3) a post-test using physical models as described below, and 4) a which-N post-test using the written symbols.

Which-N Pre- and Post-test

These tests are similar to those used in Experiments 1 and 2. There were 16 trials (**Table 1**) that sampled from one to three digit numbers to form a variety of different comparisons (e.g., transpositions, 2- vs. 3-digit numbers). All other aspects were identical to experiments 1 and 2.

Models Post-tests

The Models post-tests consisted of two tasks. The **Make-a-model** task presented children with a spoken number name and asked them to make that number using the corresponding models that they were trained with during the mapping task. In other words, this was the same task as the mapping task; only the experimenter did not make a representation of the number using the model for children to copy. There were 6 trials using numbers not used in the mapping task (**Table 1**). The second task, **Choose-a-model**, was structured just like the Which-N task—in which children were given a spoken number name and asked to choose the named number between two written numbers—only in the Choose-a-model task, they chose between two already constructed models of the number. That is, children chose between two photographs (28 cm by 30 cm) of already constructed abacus or loose discs (corresponding to the model that they used during the mapping experience) on the table—one correct and one incorrect. This task included 16 trials using numbers not used in the mapping task (**Table 1**).

Mapping Tasks

The **Standard abacus** condition (same-sized discs) was the same as the one used in Experiment 1; the diameter of the discs was 3 cm. The **Sized Abacus** condition with different-sized discs had the same overall dimension as the typical abacus, except that the discs were of different sizes: the hundred place discs had a diameter of 4 cm, the decade place discs had a diameter of 3 cm, and the unit place discs had a diameter of 2 cm. For the **Discs only** condition, discs were the same as the Sized Abacus condition but were presented loose. The set of mapping numbers for the 15 mapping trials were: 15, 186, 2, 24, 309, 38, 4, 7, 74, 851, 9, 6, 662, 50, 941. There were two orders of the mapping number sequence to which the different conditions were counterbalanced.

The procedure with the Standard abacus was the same as in Experiment 1 with two exceptions: First, there were no written numbers displayed during the mapping trials. Instead, the child was presented with the heard name, the experimenter made a model and repeated the name, and the child was encouraged to make the model. Second, in the traditional use of the abacus, discs are pushed from the bottom to the top to represent a number. Some children in Experiment 1 wanted to push the discs down not up. So, in this experiment, the represented counts were made by pushing the discs down. The discs for each count were counted by the experimenter and named as they were moved. The procedure for the Sized abacus condition was identical to the Standard abacus condition. The Discs only condition was also the same except that the number of different discs were counted and laid on the table from left to right for the hundreds, tens, and ones. As shown in **Figure 5**, within each count unit, individual discs were laid out separately (as opposed to being piled on top of each other) but spatially organized within its unit group. This spatial layout was—by design—different from how discs are organized on an abacus and was hypothesized to be more intuitive for children (e.g., without the remaining, potentially distracting discs on the vertical frame of an abacus).

Results

Children in all three conditions were successful in copying the experimenter's physical model of a spoken number, performing at 94% overall (SE = 0.9%) with no significant difference among conditions. A logistic mixed effect model, in which condition was entered as a fixed effect, participants and test items were entered as random effects, failed to detect a significant main effect of condition, χ^2 (2) = 0.64, p = 0.73. Children in the three mapping conditions also did not differ in pre-test performance on the which-N task as revealed by a logistic mixed effect model in which condition was entered as a fixed variable, participant and test item were entered as random variables, χ^2 (2) = 0.45, p = 0.80 (Standard abacus: Mean = 79%, SE = 5%, Sized abacus: Mean = 83%, SE = 5%, Discs only: Mean = 86%, SE = 4%). Again, as shown in **Figure 5**, there were considerable individual differences in children's performances at pre-test with some children performing quite poorly but others near perfectly. We will return to this fact, evident, in all three experiments in the General Discussion.

Model Post-tests

Did the mapping experience, making model representations of the spoken numbers, enable children to make those representations on their own with new numbers? In the Make-a-model post-test, children were most successful in the Discs only condition which involved choosing the right number of different sized discs and laying them on the table. A logistic mixed effect model, in which condition was entered as a fixed effect, participants and items were entered as random effects, detected a significant main effect of condition, χ^2 (2) = 5.97, p = 0.05. The mean proportion correct in the Standard abacus condition was 0.42 (SE = 0.08) and was significantly worse than those in the Discs only condition (Mean = 0.67, SE = 0.07), t (36) = 2.37, p = 0.02, or the Sized abacus condition (Mean = 0.65, SE = 0.08) with a trending significant difference, t (36) = 2.05, p = 0.047. This result suggests that one source of challenge in the use of traditional abacus is that using places on the abacus to indicate relative magnitudes (e.g., hundred, decade, unit) is not intuitive for young children. Notice, according to the Structure Mapping theory (Gentner, 1983; Gentner, 2010), there is little to help children align the components of the heard number name to the places on the traditional abacus. Using discs of different sizes at the different positions may help children align the components and thus discover the relational pattern.

The Choose-a-model post-test only required children to recognize the correct model representation of named numbers from the corresponding abacus or loose discs that children received during mapping. A logistic mixed effect model was performed in which condition was entered as a fixed effect, participants and items were entered as random effects. The model failed to detect any significant main effect of condition, χ^2 (2) = 0.79, p = 0.67 (Standard abacus: Mean = 0.68, SE = 0.05, Sized abacus: Mean = 0.71, SE = 0.04, Discs only: Mean = 0.74, SE = 0.04). In brief, children in each condition apparently learned enough to recognize (above chance) the correct model representation.

Which-N Task: Mapping Names to Written Symbols

In this study, children were never exposed to written multi-digit numbers but instead mapped number names to physical models of those quantities. Does this experience bolster children's understanding of how spoken number names map to the places of written notation? In all three logistic models, one for each condition, test time (pre, post) was entered as the fixed variable, participant and test item were entered as random variables, while the dependent variable was the accuracy of the individual test trials (i.e., 0 or 1). The Discs only condition, but not the other two conditions, led to improved performance in mapping heard number names to their written form. For the Discs only condition, there was a significant main effect of test time, χ^2 (1) = 5.02, p = 0.025; children performed significantly better at post-test (Mean = 0.89, SE = 0.03) compared to how they performed at pre-test (Mean = 0.84, SE = 0.04), indicating that mapping names to counts of different sized discs generalized to mapping novel multidigit numbers names to their written forms. For both the Standard abacus and Sized abacus, the models failed to detect any significant main effect of time (ps > 0.21). **Figure 5** shows the density plots of children's performance from pre- to post-test in the three conditions, separated by those with high and low prior knowledge. As can be seen, only the Discs only condition showed systematic improvement.

Discussion

These findings make two contributions: First, physical models for discovering how number names map to the places of written multi-digit notation do not work if the relations to be discovered are not sufficiently obvious in the model. If the models are too complex and intricate, their value as a revealing analogy—that simplifies and brings to front a main idea—is lost. The complexity of the Abacus models, and the non-obviousness of the relational structure, was clearly evident in children's difficulties in correctly creating model representations in the Models Post-test measures. Second, experiences in directly mapping number names to written notation (symbols-to-symbols mapping) are not the sole route to helping children find the relational structure; mappings to a physical model can lead to generalization and better insights about written notation if—as the Discs only condition—the analogy foregrounds the single concept to be discovered. All in all, the results of the three experiments suggest an incremental approach that does not try to do too much all at once. Within such an approach, physical models might better be used as analogies that distill a complex idea into an immediately understandable concept rather than a physical grounding of the meaning. Physical models that are not readily understandable in and of themselves cannot do that.

GENERAL DISCUSSION

Like many other domains of knowledge, place value principles of multi-digit numbers are acquired incrementally. Studies of children's informal learning about multi-digit numbers before school indicate a potentially key starting point: On their own, many children start by learning that there are different places in written multi-digit numbers and that these places signify different relative magnitudes (Byrge et al., 2014; Mix et al., 2014; Mix et al., 2017; Yuan et al., 2019). Although this entry learning does not include the multiplicative hierarchy of base-10 system, it contains the core idea of places that represent different magnitudes and strongly predicts later learning of the precise algebraic relations of multi-digit numbers (Mix et al., Under review). These previous findings—as well as the large individual differences observed in pre-test performance in the present study—indicate that children differ substantially in whether they formed this entry knowledge prior to formal schooling about place value. Given this predictive relation between this early knowledge and later learning, the implicated developmental pathway, and the individual differences, we believe that the introduction to multi-digit numbers and place value should be focused on building this early knowledge. The present findings provide useable information as to how this might be accomplished.

Our central hypothesis was that mathematical manipulatives work by serving as analogical bases—much like metaphors and analogies—to highlight the relational structure within a symbol system (Mix, 2010; Mix et al., 2019). If this is correct, then the effectiveness of symbolic or physical representations in teaching symbols does not lie on whether it accurately grounds the symbols to their complete perceivable meaning, but it is determined by whether the perceptual features of the teaching materials highlight the relational structure of the symbol system that the learner needs to learn at that point in the developmental pathway. Previous work on analogical mapping has repeatedly shown that learning is often better achieved when the component elements in the analogical base and to-be-learned system are alignable and highlight the underlying relational structures, and that aligning corresponding elements is disrupted by complicated and overly-rich stimuli that distract learners' attention away from relational patterns and to object-level details (Kotovsky and Gentner, 1996; Loewenstein and Gentner, 2001; Rattermann and Gentner, 1998; Uttal et al., 2008; Yuan et al., 2017). Under this "symbol-grounding as analogy" framework, the use of base-10 blocks in the context of first-grade place value teaching presents many challenges that may significantly limit how much students can learn. These challenges—visualized in **Figure 2**—include the large number of mappings among the different representations of written numbers, spoken number names, and base-10 blocks; the not-so-obvious perceptual structure among units of base-10 blocks (e.g., big squares, bars, small cues); and the inclusion of individual units (e.g., the 100 small squares within one big square) that may draw learners' attention to the counts of units rather than the initial understanding of the relative magnitudes of different places (e.g., hundred > decade > unit). Consistent with these considerations, neither of the two traditional mathematical manipulatives—base-10 blocks or abacus—was very effective in initially introducing students to the multi-digit number symbol system. In contrast, and as predicted by the Structure Mapping theory (Gentner, 1983; Gentner, 2010), shapes with simple but easily understood structure—big, medium, small discs—may be a more effective analogical basis for students to acquire the initial learning about places and their relative magnitudes. Base-10 blocks or abacuses may be useful tools for later learning of the precise multiplicative hierarchy of base-10 symbols—learning that goes beyond the early multi-digit number knowledge examined and measured (with the which-N and which-More tasks) in the current study—and may require substantial initial learning about the physical representations in and of themselves (e.g., their physical attributes, correspondences to symbols, and their meanings).

In contrast to the differential learning outcomes following physical models in the current studies, experience based on mapping the two symbolic forms—multi-digit number names and their corresponding written forms—have consistently produced significant improvement in children's early familiarity with multi-digit number names and their relative magnitudes. Under the "symbol-grounding as analogy" framework, there is no fundamental difference between physical models or symbolic representations in teaching symbol systems: The format of teaching materials may be different, but the key is still in aligning the elements that are in the same relation to other elements in the pattern. From a straight-up information processing perspective, heard number names and written symbols might even be better than physical representations, because the former is often more perceptually sparse and devoid of the many rich details that often characterize physical representations—and thus more "relational-orientated". It is also likely that most children already learned names for digits from 1 to 9. Their knowledge of the relative magnitudes indicated by these written symbols, their names and perhaps some knowledge of the relative magnitudes of "thousand", "hundred", and names with "-ty" (Lyons et al., 2018; Litkowski et al., 2020) enable them to align the corresponding elements in number names and written forms. This familiarity with number names may be a critical pre-requisite to generalizable learning from mappings of number names to written forms and another indicator of the importance of early parents' talks about numbers during the preschool years (Levine et al., 2010).

In the present study, both forms of symbolic representations used here—expanded cards (e.g., making 325 with 300, 20, and 5) and digits cards (e.g., making 325 with 3, 2, 5)—turned out to be effective. At first glance, mapping "three hundred" to "300" rather than to "3" seems to be a more transparent mapping. But two factors may explain the lack of difference between the two presentation formats. First, in the case of the digits card condition, one does not just map "three hundred" to "3", but to "3XX"; in other words, the spatial information—the location of each place value unit—is already baked into the mapping, and the ambiguity with respect to "where does the word hundred go with" is greatly reduced. Second, even if a learner is initially unsure about the mapping between "hundred" and "3", such mappings may become clear with repeated exposures via cross-situational statistical learning (Yu and Smith, 2007; Lany and Saffran, 2013; Rebuschat et al., 2021). For example, by encountering a series of pairs, such as "325"—"three hundred twenty five", "35"—"thirty five", "105"—"one hundred and five", the learner may accumulate enough co-occurrence information—e.g., "hundred" is often co-occur with three-digit numbers, and in such cases often occur right after the name of the leftmost digit—to arrive at the correct mapping. Examining how relational mappings are established in the midst of the ambiguity, that often characterize real-world learning data, constitutes an open and exciting future direction for building entry-level knowledge and skills about place value.

IMPLICATIONS FOR EDUCATION AND TEACHING

The current finding has several implications for introducing children to the symbolic number system and its place value principles. Foremost, children's later success in mastering place value may benefit from exposure to multi-digit numbers during the preschool and kindergarten years. Accumulating evidence with large and nationally representative samples (Byrge et al., 2014; Mix et al., 2014; Yuan et al., 2019; Yuan et al., 2020) indicates that many young children, before formal schooling, are building partial (not perfect but often correct) knowledge about how relative magnitudes are represented by place and the naming conventions for 3- and 4-digit numbers, and that this early knowledge is a strong predictor to later school learning of mathematics (Mix et al., Under review). The present result further suggests that early familiarity with these structures is learnable by preschool children with just a few trials—through mapping number names to their written forms, and/or to simple physical analogies about relative magnitudes—embedded in a game-like context without explicit teaching of the precise base-10 principles (Yuan et al., 2020).

This result may seem surprising to many researchers and educators given the well-documented difficulties that school-age children have (Baroody, 1990; Fuson et al., 1997). But these findings also make sense given what is known about young children's prodigious ability to extract, learn, and use syntactical structures by engaging in mechanisms of relational and statistical learning (Smith and Yu, 2008; Gentner, 2010). There are existing curricula (e.g., Montessori, 1917; Parrish, 2014; Mix et al., 2019) focused on teaching the relational structure of the symbolic number system, but these usually focus on older children (e.g., 5-year-old and beyond). As a result, children's early approximate knowledge about multi-digit numbers often remains a hidden competency to parents and teachers. Those children whose early experiences did not include exposure to multi-digit numbers have a hidden deficit relative to what might be critical entry knowledge. This fact—that early familiarity with written and spoken multi-digit number symbols is not only learnable, but individual differences in that understanding is highly correlated with later math learning—strongly suggests the benefit of early exposure for all children. The tasks used here—brief and easy (coached imitation)—appear suitable for early exposure.

Many researchers and educators have suggested that multi-digit number words are initial barriers to understanding because of inconsistencies (e.g., the teens numbers in English or the not initially obvious mappings of "twenty" to "two") (Miura and Okamoto, 1989; Fuson and Kwon, 1991). The present findings suggest that multi-digit number words are useful tools for entry to the symbolic number system despite the inconsistencies. The relation between language and thought is a complex one with a long history of back-and-forth debates among different theorists (Gentner and Goldin-Meadow, 2003). In the context of multi-digit number words and number learning, one often cited and emphasized result is that different language systems have different structures with some more consistent than others. One notable example is the claim that Chinese children's accelerated number learning ability—compared to English speaking children—results from the transparency of the Chinese multi-digit number words (e.g., twenty is named as two tens); in comparison, the English number words filled with inconsistencies, such as numbers in the teens range, can lead to many difficulties and represent a barrier to English speaking children's numerical development (Fuson and Kwon, 1991; Miller et al., 2000). Perhaps, because of this perception, multi-digit number words are often not the focus of formal teaching. The current study, together with a large literature from cognitive psychology (Chang et al., 2006; Lupyan, 2012), suggests that the syntactic structures of natural languages play a vital role in organizing perceptual input—including written number symbols—yielding deep latent knowledge about more abstract ideas despite the exceptions and idiosyncrasies (Yuan et al., 2020). Without experiences with number words and their inherent regularities, a child may represent a written number such as "123" as a compilation of three digits— "1", "2", and "3". But by mapping the corresponding number name "one hundred twenty-three" to the written form "123", the child may start to learn about important structural regularities, such as number words name written digits from left to right, that 3-digit numbers all have the word "hundred" in them and often appear early in the number word—regularities that are part of the place value system. This symbols-to-symbols mapping merits renewed interest by researchers as a potential critical early pathway into place value notation.

The present findings also indicate that the use of traditional manipulatives may sometimes be more of a problem than a benefit, and this may be especially the case for entry-level learning. The ultimate goal of education is for students to successfully interpret and manipulate symbols (e.g., numbers, words, equations) on the basis of their relations. In this sense, mathematical manipulatives are training wheels for learning mathematical symbols and concepts, and—like training wheels—need at some point to be abandoned. While most current educational practices in introducing students to the multi-digit number system and place value focus on "grounding" symbols into mathematical manipulatives such as base-10 blocks, studies have questioned the effectiveness of math manipulatives due to extraneous features—features that are not critical to the to-be-learned relational structure—that may be distracting to learning (McNeil et al., 2009; Kaminski and Sloutsky, 2013). But the 100 small cubes within a big base-10 block are not extraneous features; they represent the critical multiplicative relation that 100 is 100 sets of 1. Instead, manipulatives such as base-10 blocks may be ineffective in first *introducing* children to the multi-digit number system but might be useful later. Perhaps base-10 blocks should be simplified at first—that is, presented as different sized but same shaped blocks with no markings to indicate internal units—and then add the details relevant to the multiplicative relations between places. Traditional manipulatives have been shown useful for teaching of the precise base-10 relations in older children (Carbonneau et al., 2013). Another approach might be to incorporate both number symbol mappings and manipulatives and introduce the mappings in a balanced way (Mix et al., 2017). If learning the place value system is incremental, then the use of physical analogies to highlight its learning may need to be incrementally organized as well.

IMPLICATIONS FOR THE "SYMBOL-GROUNDING" PROBLEM

The idea of "symbol-grounding" has many different and nuanced interpretations (see De Vega et al. (2008) for a comprehensive review). Nonetheless, much of the theoretical discussion divides into two broad camps: symbols alone are sufficient (Pylyshyn, 1980; Fodor, 1983; Landauer and Dumais, 1997), or that perceptual and sensorimotor experiences support symbol acquisition (Barsalou, 1999, Barsalou, 2008). This dichotomy is likely to be too simple to be useful in education. At some points in learning, physical models and manipulatives—if they fit the needs of the specific task—can support learning. Physical symbols—letters and numbers—can also be used as models and with active engagement. The key question in education is when, in what way, and for what specific incremental bit of learning. Here we believe that the broad contributions of research on analogy and Structure Mapping Theory (Gentner, 2010) may help the field find useable principles.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Indiana University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

LS, KM, and RP contributed to the conception and design of the study. LY performed the statistical analyses. LY wrote the first draft of the manuscript. LS and KM contributed to the revision.

FUNDING

This research was supported by IES R305A080287 grant to LS and KM, and NIH F32 HD090827-02 grant to LY. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences or the National Institutes of Health.

ACKNOWLEDGMENTS

We thank Cody Stitzel for assisting with data collection.

REFERENCES

Alibali, M. W., and Nathan, M. J. (2012). Embodiment in Mathematics Teaching and Learning: Evidence from Learners' and Teachers' Gestures. *J. Learn. Sci.* 21 (2), 247–286. doi:10.1080/10508406.2011.611446

Ball, D. L. (1992). Magical Hopes: Manipulatives and the Reform of Math Education. *Am. Educator* 16, 14–20.

Baroody, A. J. (1990). How and When Should Place-Value Concepts and Skills Be Taught? *J. Res. Maths. Educ.* 21 (4), 281–286. doi:10.2307/749526

Barsalou, L. W. (2008). Grounded Cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (1999). Perceptual Symbol Systems. *Behav. Brain Sci.* 22 (4), 577–660. doi:10.1017/S0140525X99002149

Bartolini Bussi, M. G. (2011). Artefacts and Utilization Schemes in Mathematics Teacher Education: Place Value in Early Childhood Education. *J. Math. Teach. Educ* 14 (2), 93–112. doi:10.1007/s10857-011-9171-2

Boyer, T. W., Levine, S. C., and Huttenlocher, J. (2008). Development of Proportional Reasoning: where Young Children Go Wrong. *Dev. Psychol.* 44 (5), 1478–1490. doi:10.1037/a0013110

Bruner, J. S. (1966). *Toward a Theory of Instruction*. Harvard University Press.

Byrge, L., Smith, L. B., and Mix, K. S. (2014). Beginnings of Place Value: How Preschoolers Write Three-Digit Numbers. *Child. Dev.* 85 (2), 437–443. doi:10.1111/cdev.12162

Carbonneau, K. J., Marley, S. C., and Selig, J. P. (2013). A Meta-Analysis of the Efficacy of Teaching Mathematics with concrete Manipulatives. *J. Educ. Psychol.* 105 (2), 380–400. doi:10.1037/a0031084

Chan, W. W. L., Au, T. K., and Tang, J. (2014). Strategic Counting: A Novel Assessment of Place-Value Understanding. *Learn. Instruction* 29, 78–94. doi:10.1016/j.learninstruc.2013.09.001

Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Rev.* 113, 234–272. doi:10.1037/0033-295X.113.2.234

Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cogn. Sci.* 5 (2), 121–152. doi:10.1207/s15516709cog0502_2

Chi, M. T. H. (1978). "Knowledge Structures and Memory Development," in *Children's Thinking: What Develops?*. Editor R. S. Siegler (Psychology press), 73–96.

De Vega, M., Graesser, A. C., and Glenberg, A. M. (2008). "Reflecting on the Debate," in *Symbols and Embodiment: Debates on Meaning and Cognition*, 397–440. doi:10.1093/acprof:oso/9780199217274.003.0019

Dove, G. (2016). Three Symbol Ungrounding Problems: Abstract Concepts and the Future of Embodied Cognition. *Psychon. Bull. Rev.* 23 (4), 1109–1121. doi:10.3758/s13423-015-0825-4

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge: The MLT.

Fuson, K. C., and Briars, D. J. (1990). Using a Base-Ten Blocks Learning/Teaching Approach for First- and Second-Grade Place-Value and Multidigit Addition and Subtraction. *JRME* 21 (3), 180–206. doi:10.5951/jresematheduc.21.3.0180

Fuson, K. C. (1990). Conceptual Structures for Multiunit Numbers: Implications for Learning and Teaching Multidigit Addition, Subtraction, and Place Value. *Cogn. Instruction* 7 (4), 343–403. doi:10.1207/s1532690xci0704_4

Fuson, K. C., and Kwon, Y. (1991). Chinese-based Regular and European Irregular Systems of Number Words: The Disadvantages for English-speaking Children. *Lang. Math. Educ. Res. Pract.*, 211–226.

Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., et al. (1997). Children's Conceptual Structures for Multidigit Numbers and Methods of Multidigit Addition and Subtraction. *JRME* 28 (2), 130–162. doi:10.5951/jresematheduc.28.2.0130

Fuson, K. (1986). Roles of Representation and Verbalization in the Teaching of Multi-Digit Addition and Subtraction. *Eur. J. Psychol. Educ.* 1 (2), 35–56. doi:10.1007/BF03172568

Geary, D. C. (2007). "Development of Mathematical Understanding," in *Handbook of Child Psychology*, 777–810. doi:10.1002/9780470147658.chpsy0218

Gentner, D., Anggoro, F. K., and Klibanoff, R. S. (2011). Structure Mapping and Relational Language Support Children's Learning of Relational Categories. *Child. Dev.* 82 (4), 1173–1188. doi:10.1111/j.1467-8624.2011.01599.x

Gentner, D. (2010). Bootstrapping the Mind: Analogical Processes and Symbol Systems. *Cogn. Sci.* 34 (5), 752–775. doi:10.1111/j.1551-6709.2010.01114.x

Gentner, D., and Goldin-Meadow, S. (2003). "Whither Whorf," in *Language in Mind*. Editors D. Gentner and S. Goldin-Meadow (The MIT Press). doi:10.7551/mitpress/4117.003.0004

Gentner, D., and Stevens, A. L. (1983). *Mental Models*. Psychology Press. doi:10.4324/9781315802725

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cogn. Sci.* 7 (2), 155–170. doi:10.1207/s15516709cog0702_3

Gentner, D., and Toupin, C. (1986). Systematicity and Surface Similarity in the Development of Analogy. *Cogn. Sci.* 10 (3), 277–300. doi:10.1207/s15516709cog1003_2

Goldstone, R. L. (1998). Perceptual Learning. *Annu. Rev. Psychol.* 49, 585–612. doi:10.1146/annurev.psych.49.1.585

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D* 42, 335–346.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., and Sageman, B. (2013). Finding Faults: Analogical Comparison Supports Spatial Concept Learning in Geoscience. *Cogn. Process.* 14 (2), 175–187. doi:10.1007/s10339-013-0551-7

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., Tikoff, B., et al. (2010). Commentary: Analogical Thinking in Geoscience Education. *J. Geosci. Educ.* 58 (1), 2–13. doi:10.5408/1.3544291

Kaminski, J. A., and Sloutsky, V. M. (2013). Extraneous Perceptual Information Interferes with Children's Acquisition of Mathematical Knowledge. *J. Educ. Psychol.* 105 (2), 351–363. doi:10.1037/a0031040

Kotovsky, L., and Gentner, D. (1996). Comparison and Categorization in the Development of Relational Similarity. *Child. Develop.* 67 (6), 2797–2822. doi:10.1111/j.1467-8624.1996.tb01889.x

Landauer, T. K., and Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychol. Rev.* 104 (2), 211–240. doi:10.1037/0033-295X.104.2.211

Lany, J., and Saffran, J. R. (2013). Statistical Learning Mechanisms in Infancy. *Neural Circuit Development and Function in the Heathy and Diseased Brain* 3, 231–248. doi:10.1016/B978-0-12-397267-5.00034-0

Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., and Gunderson, E. A. (2010). What Counts in the Development of Young Children's Number Knowledge? *Dev. Psychol.* 46 (5), 1309–1319. doi:10.1037/a0019671

Litkowski, E. C., Duncan, R. J., Logan, J. A. R., Purpura, D. J., and Litkowski, E. C. (2020). When Do Preschoolers Learn Specific Mathematics Skills? Mapping the Development of Early Numeracy Knowledge. *J. Exp. Child. Psychol.* 195, 104846. doi:10.1016/j.jecp.2020.104846

Loewenstein, J., and Gentner, D. (2001). Spatial Mapping in Preschoolers: Close Comparisons Facilitate Far Mappings. *J. Cogn. Develop.* 2 (2), 189–219. doi:10.1207/S15327647JCD0202_4

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psychol.* 3, 54. doi:10.3389/fpsyg.2012.00054

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 (3), 440–457. doi:10.1037/dev0000445

McNeil, N. M., Uttal, D. H., Jarvin, L., and Sternberg, R. J. (2009). Should You Show Me the Money? Concrete Objects Both Hurt and Help Performance on Mathematics Problems. *Learn. Instruction* 19 (2), 171–184. doi:10.1016/j.learninstruc.2008.03.005

Miller, K., Major, S. M., Shu, H., and Zhang, H. (2000). Ordinal knowledge: number names and number concepts in Chinese and English. *Can. J. Exp. Psychol.* 54 (2), 129–140. doi:10.1037/h0087335

Miura, I. T., and Okamoto, Y. (1989). Comparisons of U.S. And Japanese First Graders' Cognitive Representation of Number and Understanding of Place Value. *J. Educ. Psychol.* 81 (1), 109–114. doi:10.1037/0022-0663.81.1.109

Mix, K. S., Prather, R. W., Smith, L. B., and Stockton, J. D. (2014). Young Children's Interpretation of Multidigit Number Names: from Emerging Competence to Mastery. *Child. Dev.* 85 (3), 1306–1319. doi:10.1111/cdev.12197

Mix, K. S., Bower, C. A., Hancock, G., Yuan, L., and Smith, L. B (Under review). The Development of Place Value Concepts: Principles after Approximation.

Mix, K. S., Smith, L. B., and Crespo, S. (2019). "Leveraging Relational Learning Mechanisms to Improve Place Value Instruction," in *Constructing Number - Merging Perspectives from Psychology and Mathematics Education* (Cham: Springer Publishing), 87–121. doi:10.1007/978-3-030-00491-0_5

Mix, K. S., Smith, L. B., Stockton, J. D., Cheng, Y.-L., and Barterian, J. A. (2017). Grounding the Symbols for Place Value: Evidence from Training and Long-Term Exposure to Base-10 Models. *J. Cogn. Develop.* 18 (1), 129–151. doi:10.1080/15248372.2016.1180296

Mix, K. S. (2009). "Spatial Tools for Mathematical Thought," in *The Spatial Foundations of Language and Cognition*. Editors K. S. Mix, L. B. Smith, and M. Gasser (Oxford University Press), 41–66. doi:10.1093/acprof:oso/9780199553242.003.0003

Montessori, M. (1917). *The Advanced Montessori Method A0*. New York: Frederick A. Stokes.

Namy, L. L., and Gentner, D. (2002). Making a Silk Purse Out of Two Sow's Ears: Young Children's Use of Comparison in Category Learning. *J. Exp. Psychol. Gen.* 131 (1), 5–15. doi:10.1037/0096-3445.131.1.5

O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., and Plass, J. L. (2014). Learning from Multiple Representations: An Examination of Fixation Patterns in a Science Simulation. *Comput. Hum. Behav.* 35, 234–242. doi:10.1016/j.chb.2014.02.040

Paik, J. H., and Mix, K. S. (2008). It's All Relative: Different Levels of Relational Similarity Used in Children's Comparisons. *Br. J. Develop. Psychol.* 26 (4), 499–505. doi:10.1348/026151007X260163

Parrish, S. (2014). *Number Talks: Helping Children Build Mental Math and Computation Strategies, Grades K-5*. Sausalito, CA: Math Solutions.

Pouw, W. T. J. L., van Gog, T., and Paas, F. (2014). An Embedded and Embodied Cognition Review of Instructional Manipulatives. *Educ. Psychol. Rev.* 26 (Issue 1), 51–72. doi:10.1007/s10648-014-9255-5

Pylyshyn, Z. W. (1980). Computation and Cognition: Issues in the Foundations of Cognitive Science. *Behav. Brain Sci.* 3 (1), 111–132. doi:10.1017/S0140525X00002053

R Core Team (2020). R: The R Project for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.r-project.org/.

Rattermann, M. J., and Gentner, D. (1998). More Evidence for a Relational Shift in the Development of Analogy: Children's Performance on a Causal-Mapping Task. *Cogn. Develop.* 13 (4), 453–478. doi:10.1016/S0885-2014(98)90003-X

Rau, M. A. (2017). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educ. Psychol. Rev.* 29 (Issue 4), 717–761. doi:10.1007/s10648-016-9365-3

Rebuschat, P., Monaghan, P., and Schoetensack, C. (2021). Learning Vocabulary and Grammar from Cross-Situational Statistics. *Cognition* 206, 104475. doi:10.1016/j.cognition.2020.104475

Richland, L. E., and Simms, N. (2015). Analogy, Higher Order Thinking, and Education. *Wiley Interdiscip. Rev. Cogn. Sci.* 6 (2), 177–192. doi:10.1002/wcs.1336

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., and Lyons, E. A. (2017). Supporting Mathematical Discussions: the Roles of Comparison and Cognitive Load. *Educ. Psychol. Rev.* 29 (1), 41–53. doi:10.1007/s10648-016-9382-2

Sarama, J., and Clements, D. H. (2009). "Concrete" Computer Manipulatives in Mathematics Education. *Child. Develop. Perspect.* 3 (3), 145–150. doi:10.1111/j.1750-8606.2009.00095.x

Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2015). Afex: Analysis of Factorial Experiments. R package version 0.13–145.

Singmann, H., and Kellen, D. (2019). "An Introduction to Mixed Models for Experimental Psychology," in *New Methods in Cognitive Psychology* (Routledge), 4–31. doi:10.4324/9780429318405-2

Smith, L. B., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106 (3), 1558–1568. doi:10.1016/j.cognition.2007.06.010

Socher, R., Ganjoo, M., Manning, C. D., and Ng, A. Y. (2013). Zero-Shot Learning Through Cross-Modal Transfer. *Advances in Neural Information Processing Systems*, 935–943.

Son, J. Y., Smith, L. B., and Goldstone, R. L. (2012b). Connecting Instances to Promote Children's Relational Reasoning. *J. Exp. Child. Psychol.* 108 (2), 260–277. doi:10.1016/j.jecp.2010.08.011.Connecting

Son, J. Y., Smith, L. B., Goldstone, R. L., and Leslie, M. (2012a). The Importance of Being Interpreted: Grounded Words and Children's Relational Reasoning. *Front. Psychol.* 3, 45. doi:10.3389/fpsyg.2012.00045

Son, J. Y., Smith, L. B., and Goldstone, R. L. (2008). Simplicity and Generalization: Short-Cutting Abstraction in Children's Object Categorizations. *Cognition* 108 (3), 626–638. doi:10.1016/j.cognition.2008.05.002

Sowell, E. J. (1989). Effects of Manipulative Materials in Mathematics Instruction. *J. Res. Maths. Educ.* 20 (5), 498. doi:10.2307/749423

Steels, L. (2008). "The Symbol Grounding Problem Has Been Solved, So What's Next?," in *Symbols and Embodiment: Debates on Meaning and Cognition* (Oxford University Press), 223–244. doi:10.1093/acprof:oso/9780199217274.003.0012

Stolz, S. A. (2015). Embodied Learning. *Educ. Philos. Theor.* 47 (5), 474–487. doi:10.1080/00131857.2013.879694

Taddeo, M., and Floridi, L. (2005). Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research. *J. Exp. Theor. Artif. Intell.* 17 (4), 419–445. doi:10.1080/09528130500284053

Thompson, C. A., and Opfer, J. E. (2010). How 15 Hundred Is like 15 Cherries: Effect of Progressive Alignment on Representational Changes in Numerical Cognition. *Child. Dev.* 81 (6), 1768–1786. doi:10.1111/j.1467-8624.2010.01509.x

Uttal, D. H., Gentner, D., Liu, L. L., and Lewis, A. R. (2008). Developmental Changes in Children's Understanding of the Similarity between Photographs and Their Referents. *Dev. Sci.* 11 (1), 156–170. doi:10.1111/j.1467-7687.2007.00660.x

Wang, W., Zheng, V. W., Yu, H., and Miao, C. (2019). A Survey of Zero-Shot Learning, *ACM Trans. Intell. Syst. Technol.* 10 (Issue 2), 1–37. doi:10.1145/3293318

Wills, T., Shipley, T. F., Chang, B. L., Cromley, J. G., and Booth, J. L. (2008). "What Gaze Data Reveal about Coordinating Multiple Mathematical Representations," in Proceedings of the Annual Meeting of the Cognitive Science Society, 3113–3118.

Yu, C., and Smith, L. B. (2007). Rapid Word Learning under Uncertainty via Cross-Situational Statistics. *Psychol. Sci.* 18 (5), 414–420. doi:10.1111/j.1467-9280.2007.01915.x

Yuan, L., Prather, R. W., Mix, K. S., and Smith, L. B. (2019). Preschoolers and Multi-Digit Numbers: A Path to Mathematics through the Symbols Themselves. *Cognition* 189, 89–104. doi:10.1016/j.cognition.2019.03.013

Yuan, L., Uttal, D., and Gentner, D. (2017). Analogical Processes in Children's Understanding of Spatial Representations. *Dev. Psychol.* 53 (6), 1098–1114. doi:10.1037/dev0000302

Yuan, L., Xiang, V., Crandall, D., and Smith, L. (2020). Learning the Generative Principles of a Symbol System from Limited Examples. *Cognition* 200, 104243. doi:10.1016/j.cognition.2020.104243

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Education

ORIGINAL RESEARCH
published: 16 September 2021
doi: 10.3389/feduc.2021.629201

Disentangling the Effects of SFON (Spontaneous Focusing on Numerosity) and Symbolic Number Skills on the Mathematical Achievement of First Graders. A Longitudinal Study

Noemi Gloor[1], Delia Leuenberger[1,2] and Elisabeth Moser Opitz[1]*

[1]*Institute of Education, University of Zurich, Zurich, Switzerland,* [2]*University of Applied Sciences in Special Needs Education, Zurich, Switzerland*

Edited by:
Ricardo Moura,
University of Brasilia, Brazil

Reviewed by:
Dor Abrahamson,
University of California, Berkeley, United States
Minna Hannula-Sormunen,
University of Turku, Finland

***Correspondence:**
Noemi Gloor
noemi.gloor@ife.uzh.ch

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

Received: *13 November 2020*
Accepted: *02 September 2021*
Published: *16 September 2021*

Citation:
Gloor N, Leuenberger D and Moser Opitz E (2021) Disentangling the Effects of SFON (Spontaneous Focusing on Numerosity) and Symbolic Number Skills on the Mathematical Achievement of First Graders. A Longitudinal Study. Front. Educ. 6:629201. doi: 10.3389/feduc.2021.629201

Research has established that Spontaneous Focusing on Numerosity (SFON) and symbolic number skills (e.g., counting out loud, counting objects, linking small magnitudes and numbers) are predictors of mathematical achievement in primary school. However, little is known about the relationship between SFON and symbolic number skills, or whether one of these factors is more influential on a child's subsequent mathematical achievement. This study investigated the effect of SFON and symbolic number skills on mathematical achievement at the end of Grade 1 by controlling for first language, gender, working memory and nonverbal IQ. Participants were N = 1,279 first graders. SFON, symbolic number skills and control variables (first language, gender, working memory, and nonverbal IQ) were measured at the beginning of Grade 1. SFON was assessed with a verbally-based task. Data on mathematical achievement was collected at the end of Grade 1. Descriptive statistics demonstrated that the children's SFON was relatively low at the beginning of Grade 1. Structural equation modeling was used to examine the relationship between SFON, symbolic number skills and mathematical achievement at the end of Grade 1. The results revealed a weakly significant correlation between SFON and symbolic number skills. SFON and symbolic number skills were significant predictors of mathematical achievement at the end of Grade 1. However, the effect of symbolic number skills on mathematical achievement was greater than the effect of SFON. It is therefore concluded that numerical skills are more important than SFON for predicting mathematical achievement over the course of first grade.

Keywords: longitudinal study, primary school, mathematical achievement, spontaneous focusing on numerosity (SFON), symbolic number skills

INTRODUCTION

Spontaneous Focusing on Numerosity (SFON) and symbolic number skills have both been identified as important predictors of mathematical achievement gain (e.g., Hannula-Sormunen et al., 2015; Gallit et al., 2018). Hannula et al. (2005) were the first to investigate SFON. SFON is defined as a "process of spontaneously [. . .] focusing attention on the exact number of a set of items or incidents" (Hannula et al., 2010, p. 395). Young children pay attention to quantitative aspects of their environment: They count steps when climbing the stairs, compare the number of cookies they would like to eat or the number of objects in a storybook. The term "spontaneous" means that the process of focusing on numbers is self-initiated and not guided by others. "This attentional process is needed for triggering exact number recognition processes and using the recognized exact number in action because exact number recognition is not a totally automatic process that would take place every time a person faces something to enumerate" (Hannula et al., 2010, p. 395). Hannula-Sormunen et al. (2020) emphasize that this process of focusing attention on the exact number of objects in their surroundings is a skill that children have to learn. It enables them to efficiently utilize the innate mechanisms of subitizing for active quantification processes. A child's SFON performance reflects their tendency to focus on the numerical, so discussions of SFON often refer to SFON tendencies, but in this paper we use the term SFON on its own.

SFON seems to be important for later mathematical achievement. Empirical findings reveal a relationship between children's SFON and their early mathematical skills (e.g., Hannula et al., 2005; Hannula and Lehtinen, 2005; Hannula et al., 2007; Edens and Potter, 2013). There is also evidence that a child's SFON is related to their subsequent mathematical achievement in primary school (e.g., Hannula et al., 2010; Hannula-Sormunen et al., 2015; McMullen et al., 2015). But some studies suggest number skills acquired before starting school are also important predictors of mathematical achievement in primary school (e.g., Jordan et al., 2007; Krajewski and Schneider, 2009; Conoyer et al., 2016; Gallit et al., 2018). These skills–variously termed "early numeracy" (Conoyer et al., 2016), "number sense" [Jordan et al. (2007), or "early quantity-number competencies" (Krajewski and Schneider, 2009)–include the ability to count out loud, compare numbers and magnitudes, link small magnitudes and numbers, and do simple calculations. Number skills can be divided into symbolic and non-symbolic categories. Symbolic skills (e.g., linking numbers and magnitudes, reading numbers] have been shown to be especially important for the development of further mathematical skills (e.g., Kolkman et al., 2013; Göbel et al., 2014). So both SFON and symbolic number skills are important for mathematical learning although little is known about the relationship between them, or whether one has a greater effect on subsequent mathematical achievement than the other. This study investigates how a child's SFON and symbolic number skills at the beginning of Grade 1 may relate to mathematical achievement at the end of Grade 1.

Measuring SFON

There are a variety of tasks designed to measure SFON, and studies show that the type of task can have an influence on measured SFON (Batchelor et al., 2015; Rathé et al., 2016; Nanu et al., 2020). According to Hannula (2005), the following criteria are important when assessing SFON. In order to avoid "numerical hints," only tasks which are new to the children and tasks without mathematical context should be used. Furthermore, the use of mathematical vocabulary (e.g., count, number word) should be avoided both before and during the test. SFON tasks should include small numbers of objects that are easy to enumerate. Finally, SFON tasks should not exceed the children's working memory capacity or visuo-motor or verbal comprehension skills (ibid.).

Hannula et al. (e.g., Hannula et al., 2005; Hannula and Lehtinen, 2005) developed different types of action-based tasks: The imitation task, the model task, the finding task, and the selection task. In the imitation task, children are instructed to imitate the action of a test administrator (e.g., posting a certain number of blue and red envelopes into a mailbox). In the model task, the children have to carefully observe the activity of a test administrator (e.g., depicting a dinosaur with stamps) and copy the dinosaur as precisely as possible. In the finding task, the test administrator hides a toy (e.g., a gold ingot) under one of three objects (e.g., wooden hats). The children have to remember where the toy was hidden and lift the correct cover. In the selection task, the children are told to give a certain number of objects to a creature (e.g., "This creature has a problem. The creature's legs feel terribly cold. Fortunately, there are boxes of socks under the cloth. Give this creature its own box of socks." Hannula et al., 2005, p. 70). These tasks all have some limitations. The imitation task and the model task could, possibly, be successfully completed using imitation alone, without any numerical reasoning, especially when conducted with small quantities of objects. Also, the children have to focus on the specific activity presented by the administrator and their attention has to be drawn to this activity from the very beginning of the task. Therefore, the result might be affected by the children's attention capacity and/or working memory. This is not the case for the selection task, which requires numerical thinking when comparing quantities.

A different type of SFON task is the picture task, which was developed by Batchelor et al. (2015). In the picture task, the children are shown a picture with different objects in varying numbers (e.g., a river with three boats, four ducks and two trees). The children are asked to describe what they see. Contrary to the action-based task, here the focus can be on different dimensions. Children may not only focus on the number of objects (e.g., "two girls") but also on the colors of the objects (e.g., "a red shirt") or other aspects like emotions (e.g., "the girls look happy"). In addition, the picture task is quick and easy to handle, and no specific material is required. Furthermore, the scoring is simple, and no additional analyses are necessary (ibid.). The picture task, however, also has limitations. It requires active language skills like vocabulary and number words and may be challenging for second language learners or children with language impairment (ibid.). In addition, a child's answers might be affected by his or her interests. While some children are more interested in numbers, others might focus on colors or shapes, or the vocabulary (e.g., "I forgot what duck means in Spanish.").

A Literature Review of SFON

Existing studies have examined SFON in different age groups. The findings of longitudinal studies reported high stability of children's SFON across time (e.g., Hannula and Lehtinen, 2005; Bojorque et al., 2017). In the study by Hannula and Lehtinen (2005), SFON was measured at ages 4, 5 and 6 using different action-based SFON tasks. In spite of the 2-years time period and the different contexts of the tasks, there was reasonable stability in the children's SFON.

A cross-sectional study by Kucian et al. (2012) demonstrated that children with mathematical learning disabilities aged 7–11 had significantly lower SFON, as measured by two action-based tasks, than children without disabilities. These findings could not be explained by IQ, age or gender. It remains unclear, however, whether a lower SFON leads to low mathematical skills or whether mathematical difficulties lead to a lower SFON. Gray and Reeve (2016) identified preschoolers' math ability profiles and examined how number-specific markers like SFON (measured using three action-based tasks) and dot enumeration, as well as general markers (working memory, response inhibition, attention, and vocabulary), were associated with profiles. Results showed that the numerical markers were significantly associated with the math ability profiles, whereas the association between the other markers was either not significant or only marginally so.

Findings of cross-sectional studies provide empirical evidence that a child's SFON, measured using action-based tasks, is positively correlated with number sense and early mathematical skills (e.g., Hannula and Lehtinen, 2005; Edens and Potter, 2013). Hannula and Lehtinen (2005) investigated SFON and early mathematical skills in preschoolers. Results showed that the children's SFON correlated with number sequence elaboration, counting of objects, and basic arithmetic skills such as addition and subtraction. These relationships remained significant after controlling for nonverbal IQ and the comprehension of verbal instructions. Edens and Potter (2013) found that 4-year old children who spontaneously focused on numerosity had better counting skills.

Longitudinal studies have also provided insight into SFON and its relationship with mathematical skills. A study by Hannula et al. (2007) investigated how SFON, measured using action-based tasks, is related to subitizing-based enumeration and verbal and object counting skills in four and five-year-old children. Results showed that SFON was directly related to verbal counting skills even when subitizing-based enumeration was entered in the model. The association between SFON and object counting skills was mediated by subitizing-based enumeration. Further, empirical evidence shows that children's SFON is also related to subsequent mathematical achievement in primary school (e.g., Hannula et al., 2010; Hannula-Sormunen et al., 2015; McMullen et al., 2015). Hannula et al. (2010) showed that children's SFON in kindergarten, measured using an action-based task, accounted for a domain-specific and significant, but small, variance (2%) at the end of Grade 2. Children's SFON was a significant predictor of arithmetic skills at the end of Grade 2, but not of reading skills. The domain specificity of SFON is also supported by the findings of Nanu et al. (2018). In their study, SFON measured at age five with action-based tasks predicted arithmetic fluency and number line estimation in fifth grade, but not rational number knowledge or mathematical achievement. Hannula-Sormunen et al. (2015) analyzed the effect of children's SFON, again measured with action-based tasks, subitizing, and counting skills on their mathematical achievement at age of 12. Subitizing-enumeration skills were tested at age five. SFON and counting skills were then assessed a year later, at age six. Their results showed that children's SFON and counting skills were both predictors of mathematical achievement at age 12. However, after controlling for nonverbal IQ, only SFON predicted mathematical achievement. The association between subitizing and mathematical achievement was mediated by SFON and counting skills. McMullen et al. (2015) followed a sample cohort to investigate how children's SFON and counting skills as measured at age six related to their rational number conceptual knowledge 6 years later. These results suggest that SFON, measured using action-based tasks, and counting skills predict rational number conceptual knowledge, thus lending support to the hypothesis that SFON is a predictor of a child's future mathematical achievement. But it should be noted that the sample in that longitudinal study was very small (N = 36).

Chan and Mazzocco (2017) investigated children's and adults' attention to numbers, which is a concept related to SFON. The aim of the study was "to address the 'spontaneity' and the malleability of SFON [...] under varying conditions" (p. 77). The attention to numbers was measured using a picture-matching task, where the participants had to choose one of four pictures that matched a target picture. Results demonstrated that only 8–10% of children's best matches were number based, while 21% of adults' were number based. Children's attention to number did not increase when prompted to search for other matches. In addition, children's attention to number was affected by competing features (e.g., color, shape, position, or quantity).

Hannula et al. (2005) investigated the possibility of increasing SFON with an intervention. The results demonstrated that SFON, measured using action-based tasks, can be enhanced with a guided intervention in preschool that focused on numerical activities. However, this was only the case for children with high SFON at the first measurement point. Children with no or low SFON at the beginning of the study did not respond to the intervention. Another intervention study by Hannula-Sormunen et al. (2020) tested the effects of two early numeracy intervention programs on SFON and early numerical skills. The intervention programs were integrated into daily day care routines and included activities such as noticing numbers and number recognition that were aimed at developing the subitizing mechanism and paying attention to numerical aspects of everyday activities. The results showed that the intervention programs had a positive effect on children's SFON as measured by action-based tasks, from pretest to posttest and a long-term effect on cardinality-related skills from posttest to delayed posttest.

Braham et al. (2018) investigated whether SFON measured with action-based tasks could be enhanced through guided parent-child interactions in a children's museum. Children whose parents had received the numerical intervention program showed higher SFON scores than children whose parents had not. These findings suggest that parents can foster children's SFON using numerical prompts in informal play settings.

In conclusion, this research overview shows that SFON is related to a variety of other mathematical skills, in both the short and long term. Other research, presented below, emphasizes the significance of the construct of number skills, especially symbolic number skills, for mathematical achievement gain.

Symbolic Number Skills and Its Significance for Mathematical Competence

Research also shows that early number skills such as number knowledge, verbal counting, object counting, and non-symbolic or numerical magnitude comparison are strong predictors for later mathematical achievement in primary school (e.g., Jordan et al., 2007; Krajewski and Schneider, 2009; Conoyer et al., 2016; Toll et al., 2016; Gallit et al., 2018). The process of acquiring number skills includes the innate ability to recognize a small number of items without counting called "subitizing" (e.g., Wynn, 1995), as well as skills that must be acquired through social mediation and education, like counting competence or writing numbers (Dehaene, 2001; Dowker, 2005; Kolkman et al., 2013; Hannula-Sormunen et al., 2020). Frydman (1995) and Simon and Vaishnavi (1996) stress that the process of subitizing contains non-numerical knowledge and therefore differs from other mathematical learning processes like counting. The mediation of numerical skills begins in early childhood when children mimic number sequences used by their parents or siblings (e.g., Fuson, 1988).

Krajewski (2003) found that early numerical skills, as measured 6 months before the start of school, made the greatest contribution to the prediction of arithmetic performance in first grade. Results of this study revealed that poor early numerical skills could predict mathematical difficulties at the end of first and second grade much better than measures of intelligence. Jordan et al. (2007) investigated number sense and its development as predictors for formal math achievement in first grade. They reported that number sense performance in kindergarten and number sense growth accounted for 66% of the variance in math achievement in first grade. Dornheim (2008) found that the early numerical skills in kindergarten were the main predictor of arithmetic achievement in first and second grade. This was also confirmed by the study of Gallit et al. (2018).

There is, however, empirical evidence that some aspects of number skills might be more important than others. According to Krajewski and Schneider (2009), the linkage of quantities and numbers represents the most important concept for successful mathematical learning in primary school. More recent research provides evidence that the differentiation between non-symbolic number skills (e.g. comparing magnitudes) and symbolic number skills (e.g. numerical tasks like counting and Arabic symbols) (e.g., Kolkman et al., 2013) seems to be crucial (e.g., Missall et al., 2012; Kolkman et al., 2013; Göbel et al., 2014; Toll et al., 2016; Caviola et al., 2020). Kolkman et al. (2013) investigated the role of non-symbolic and symbolic skills in early numerical development with children at age 4, 5 and 6. Their results provided evidence for the predominant role of symbolic skills as compared to non-symbolic skills in the development of mapping skills (linkage of number symbols and their corresponding quantities). According to Missall et al. (2012), symbolic skills (comparing numbers, inserting a missing number in a number sequence) are the most robust factors for predicting later math performance. They examined predictive relationships from kindergarten through third grade. Göbel et al. (2014) analysed the impact of symbolic knowledge of the Arabic numeral system and magnitude comparison on arithmetical skills 11 months later in a sample of first graders. Path models revealed that knowledge of the Arabic numeral system predicted an increase in arithmetic skills, whereas magnitude comparison skills had no impact. Caviola et al. (2020) reported similar results from a sample of second graders. Non-symbolic magnitude comparison had no association with mathematical performance. Toll et al. (2016) found that symbolic number skills (which they term number sense) measured at the end of the first year of kindergarten (M_{age} = 4.96) are the strongest predictors of mathematical performance (math facts and math problems) in first grade. Non-symbolic number sense (dot comparison) was only a predictor of problem solving ability. The importance of symbolic number skills for later arithmetic skills was confirmed in a review paper by Szkudlarek and Brannon (2017). Symbolic numerical competence also plays an important role in interventions aiming to improve SFON. According to Hannula-Sormunen et al. (2020) and Braham et al. (2018), symbolic numerical activities resulted in SFON achievement gains.

Non-numerical Predictors of Mathematical Competence

SFON and number skills, as well as mathematical achievement gain, are influenced by first language, gender, working memory and nonverbal IQ. Anders et al. (2012) found that the first language of the parents had an impact on the number skills of young children and their achievement gain. Also, studies by Kuratli Geeler (2019) and Sale et al. (2018) revealed that having language of classroom education as a first language had a significant influence on the numerical competence of children in kindergarten. In addition, controlling for first language is crucial when a picture-based task is used to assess SFON. The relationship between gender and numerical competence is still unclear. Some studies found no differences in numerical competence between boys and girls (e.g., Dornheim, 2008; Niklas and Schneider, 2012; Sale et al., 2018), while the research of Kuratli Geeler (2019) revealed higher numerical competences for boys in kindergarten, especially in tasks which required symbolic representation. Anders et al. (2012), on the other hand, showed that girls have a higher numerical competence in preschool. For Grade 1, the picture is more consistent, and several researchers reported higher achievement levels for boys (Krajewski, 2003; Niklas and Schneider, 2012; Sale et al., 2018; Kuratli Geeler, 2019). There is also evidence that working memory is significantly related to mathematical achievement (e.g., De Smedt et al., 2009) and therefore important to control for. Batchelor et al. (2015) found no significant correlation between SFON and working memory. Furthermore, previous study results reveal that nonverbal IQ has an impact on number skills and mathematical achievement (e.g., Krajewski and Schneider, 2009; Kuratli Geeler, 2019), whereas there was no relationship between nonverbal IQ and SFON measured with an imitation task (Hannula et al., 2010).

A review of the literature reveals that existing studies highlight the significance of SFON and identify SFON as a predictor of subsequent mathematical performance. Nevertheless, there are some research gaps. First, there is empirical evidence that symbolic number skills are also important predictors of mathematical achievement. Researchers have examined the relationship between SFON and certain aspects of mathematical competence, such as counting or subitizing, and the effect of SFON and counting competence on later mathematical achievement, but they have not yet looked at whether or how SFON and a range of symbolic numerical competences present at the beginning of school career can affect later mathematical achievement gain. Second, longitudinal studies analyzing the effect of SFON on mathematical achievement have often been conducted using small study samples. Third, the research overview shows that most studies have investigated the effect of children's SFON on mathematical achievement using action-based tasks (e.g., Hannula and Lehtinen, 2005; Hannula et al., 2007; McMullen et al., 2015; Nanu et al., 2018). The picture task has only been used in a few instances (e.g., Batchelor et al., 2015; Rathé et al., 2019). Finally, even if most studies on SFON included non-mathematical predictors, studies that have a broad range of control variables are rare. This study aims to close these research gaps by investigating the effect of SFON and symbolic number skills on mathematical achievement gain using a verbally-based SFON task and a large sample of 1,279 first graders. First language, gender, nonverbal IQ, and working memory are included as control variables.

The following research questions are addressed:

1) To what extent do children spontaneously focus on numerosity at the beginning of Grade 1? Results from previous studies (e.g. Hannula and Lehtinen, 2005; Hannula et al., 2010) indicate that there will be a large variance in children's SFON.
2) Is there a relationship between SFON and symbolic number skills? Research by Hannula and Lehtinen (2005), and Edens and Potter (2013), suggest there will be a moderately significant correlation.
3) Do SFON and symbolic number skills have an effect on mathematical achievement at the end of Grade 1, controlling for first language, gender, working memory and nonverbal IQ? On the basis of previous research, it is hypothesized that SFON and symbolic number skills both have an effect on mathematical achievement at the end of Grade 1.
4) Is SFON or symbolic number skills more important for mathematical achievement at the end of Grade 1? It is assumed that symbolic number skills have a greater effect on mathematical achievement gain at the end of Grade 1 than does SFON.

TABLE 1 | Descriptive characteristics of the sample.

	n (%)
Pupils	1,279
Gender	
Male	651 (50.9)
Female	628 (49.1)
First language	
German	570 (44.6)
German and other	316 (24.7)
Other	244 (19.1)
Missing	149 (11.6)

First language, gender, nonverbal IQ, and working memory are included as control variables. SFON is assessed with a picture task that is verbally-based. Therefore, it is likely that first language might influence a child's SFON. Results from the studies presented in the literature review suggest that gender–boys performing better, on average–and nonverbal IQ explain variance in SFON, number skills and mathematical achievement at the end of Grade 1. It is expected that first language and working memory will also influence symbolic number skills and mathematical achievement at the end of Grade 1.

MATERIALS AND METHODS

Participants and Context of the Study

In the Swiss education system, 1 year of kindergarten is compulsory and kindergarten is free of charge. Therefore, all children attend at least 1 year in kindergarten. Numerical instruction following a compulsory curriculum begins in kindergarten (Deutschschweizer Erziehungsdirektoren-Konferenz (D-EDK), 2016). This numerical instruction focuses on oral counting up to 20, counting backwards and forwards from every possible number up to 10, object counting, comparing numbers, and using number words like "bigger", "smaller", "more" or "less" (ibid.).

Participants were 1,279 first graders (49.1% girls, $M_{age} = 6.82$, $SD = 0.38$) from 77 primary schools in German-speaking Switzerland (**Table 1**). Invitation letters were sent to several schools via the school authorities. Teachers decided voluntarily whether they wished to participate. All parents gave written

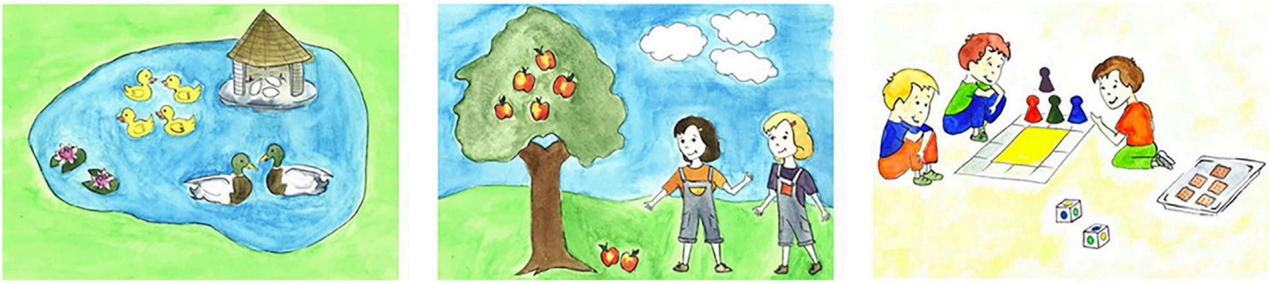


FIGURE 1 | The pictures used in the verbally-based SFON task, drawn by Luisa Leliuc.

consent for the participation of their children in the study. 578 children (46.2%) had German as a first language, 316 children (25.3%) were bilingual and 232 children (18.5%) had another first language (Missing n = 125; **Table 1**).

Data was collected over one school year. At the beginning of Grade 1 (t1), SFON, working memory and nonverbal IQ were assessed, with the children working individually with a test administrator in a quiet room at the children's school. The symbolic number skills test was conducted with groups of 8–12 children at the beginning of Grade 1 (t1), after the SFON test. The test to measure mathematical achievement in Grade 1 (t2) was carried out with the whole class at the end of the school year. The teachers completed a questionnaire on children's first language, gender, and age at the beginning of the school year.

Instruments

SFON

So that their SFON could be assessed, the children completed a picture task. The picture task was used because compared to an action-based task, it is quick and easy to handle, the scoring is simple, and no additional analyses are necessary. The pictures used in this study were variations of those used by Batchelor et al. (2015). Three pictures (**Figure 1**) were presented one after the other on a screen (13"–15"), in the same order for each child.

The test administrator introduced the SFON task as follows: "I am going to show you different pictures. We are interested in what children will tell us about these pictures. This is the first picture. What do you see in this picture?" Each single statement (e.g., yellow ducks, a pond, a pond with ducks, two girls, a T-Shirt with a flower) was scored as numerical or non-numerical. For efficiency, given the size of the sample, only the first four statements per child were written down. Each numerical answer (e.g., two ducks, three boys) was scored with 1, regardless of whether the number was correct. Answers like "some ducks" or "both girls" were scored 0. Because the German word for "a" is the same as the word for one ("a tree"), these answers were excluded. The children could achieve a maximum score of 12. Confirmatory factor analysis confirmed that the scale was unidimensional and Cronbach's alpha was 0.87.

Symbolic Number Skills

The symbolic number skills assessment involved 22 items that covered the following topics: counting objects (7 apples, 11 dots, 23 dots) and linking the result with the correct number (3 items), comparing numbers up to 20 (5 items), writing selected numbers of the number sequence up to 20 (11 items), and writing the matching mathematical term for a picture (e.g., 2 + 3 to a picture with two red and three blue balloons; 3 items). Test instructions were given by the test administrator. Some of the tasks were explained with an example. In the number comparison task, two numbers (e.g., 3 and 1) were each written in a box in the booklet. The box with the bigger number was checked. "Here are two numbers in a box, 1 and 3. 3 is more than 1, therefore, this box is checked. Here are two other numbers, each in a box (numbers 6 and 2). You have to check the box with the bigger number." Cronbach's alpha was 0.88.

Mathematical Achievement at the End of Grade 1

Mathematical achievement was tested at the end of Grade 1 using an author developed test prepared for publication. The test included 27 items. The following topics were assessed: counting by steps (completing the number sequence 3, 5, 7 ... 15 and 12, 14, 16, ... 24; 2 items), number decomposition (e.g., 20 = __ + __ + __; 3 items), doubling the numbers 4, 7, 12 (3 items), halving the numbers 16, 18, 22 (3 items), addition (7 + __ = 13, 11 + __ = 19, 18 + __ = 23, 80 + __ = 100; 4 items), subtraction (9–3, 18–8, 17–12, 14–7; 4 items), and word problems (picture of a toy with a price tag: picture with a Swiss bill: you pay with the bill; how much change do you get? 8 items). Most of the test instruction was given using tables and pictures and the test administrator was allowed to read out the short instructions. The counting by steps task was presented in the following way: 3, 5, 7, __, __, __, 15. "Look at these numbers: 3, 5, 7. The numbers continue in the same way. Which numbers fit into the gaps? Write the correct number in the gaps." Rasch analyses were conducted to assess the quality of the test. Weighted likelihood estimate (WLE) of reliability was 0.79. The item fit was acceptable (0.89–1.27) (Wilson, 2005). The variable was z-standardized.

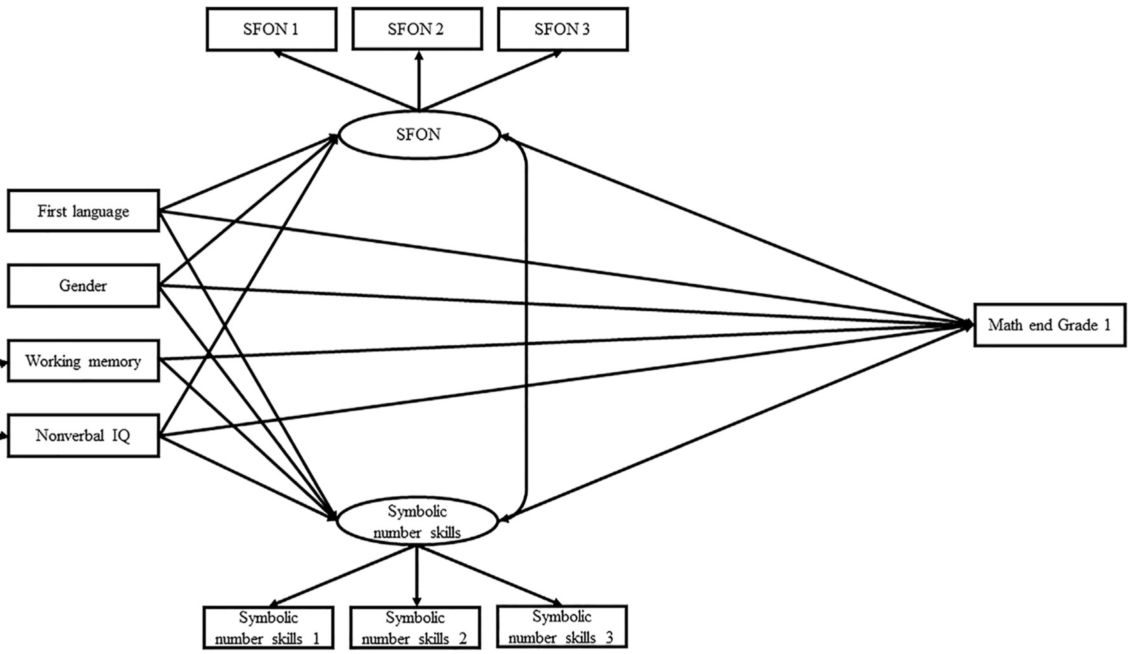


FIGURE 2 | Hypothesized Model. SFON and symbolic number skills will predict mathematical achievement at the end of Grade 1 when data are controlled for non-specific predictors (first language, gender, working memory, nonverbal IQ). The oval symbols are latent variables that represent the variance shared by multiple indicators. The square symbols represent manifest variables.

TABLE 2 | Descriptive Statistics for all measures.

	n	*M*	*SD*	Range
SFON picture 1	1,255	0.76	1.32	0–4
SFON picture 2	1,255	1.22	1.23	0–4
SFON picture 3	1,255	1.40	1.32	0–4
SFON total	1,255	3.38	3.42	0–12
Symbolic number skills	1,235	18.99	3.91	2–22
Mathematical achievement	1,130	18.56	6.42	0–27
Working memory	1,231	12.59	2.73	1–19
Nonverbal IQ	1,235	15.94	5.29	0–30

Nonverbal IQ

Nonverbal IQ was measured using two subtests of CFT 1-R (Weiß and Osterland, 2013): similarities (15 items) and matrices (15 items). Cronbach's alpha for 30 items was 0.85.

Working Memory

To measure children's working memory, the corsi blocks (10 items) and number sequence backward (10 items) subtests of a working memory test battery for children aged 5–12 (AGTB 5–12) (Hasselhorn et al., 2012) was used. Cronbach's alpha for 20 items was 0.72.

First Language

First language was assessed with a teacher questionnaire. Teachers were asked to indicate which language the children speak at home on a 3-point Likert-scale (1 = only German, 2 = bilingual; German and other language, 3 = other languages). Two different variables were calculated: A dichotomous variable with the groups German speaking only/bilingual and other language, and a second variable with the groups German speaking only and bilingual/other language. The analyses were carried out with both variables, but no difference in the results was found. Therefore, results for the groups German only and bilingual and other language are reported.

Statistical Analyses

To test whether a child's SFON and symbolic number skills have an effect on mathematical achievement at the end of Grade 1, a structural equation model was set up using the lavaan package (Rosseel, 2012) on R software version 3.5.2. SFON and symbolic number skills, as well as the non-specific variables, namely, first language, gender, working memory and nonverbal IQ, were included as predictors (**Figure 2**). Based on the results of previous studies on the relationships between SFON, symbolic number skills and mathematical achievement, it was assumed that SFON and symbolic number skills both had a direct effect on mathematical achievement at the end of Grade 1. Further, a correlation between SFON and symbolic number skills was assumed. In addition, the non-specific variables first language, gender, working memory and IQ were expected to predict SFON, symbolic number skills and mathematical achievement. As SFON was measured using the picture task, a child's SFON would not be affected by working memory. Finally, a correlation between working memory and nonverbal IQ was assumed.

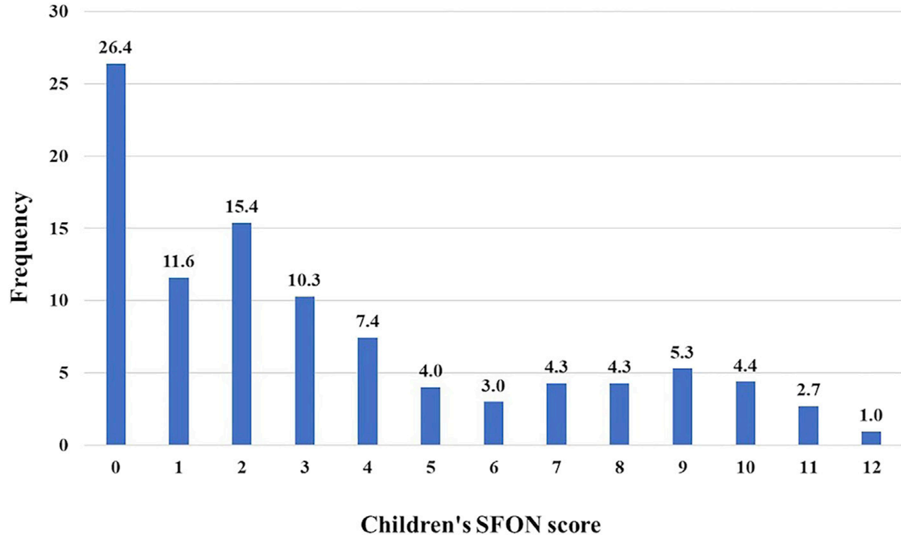


FIGURE 3 | Frequencies of children's SFON.

In the hypothesized model, SFON and symbolic number skills were included as latent variables and parcels were built. Parcels help to reduce the complexity of models. Additionally, structural equation models based on parceled data lead to more stable estimates and fit the data better (Matsunaga, 2008). Due to the unidimensionality of the SFON construct, the parcels were allocated to the task. Homogenous or heterogenous parceling strategies can be used to build parcels of multidimensional constructs. In homogenous parceling strategies, similar items are placed in the same parcel, while "in heterogenous parceling strategies, items that share a source of systematic variation are distributed across different parcels either randomly or systematically" (Marsh et al., 2013, p. 260). Because homogenous parcels lead to bad factor loadings, heterogenous parceling strategies were used for the construct of symbolic number skills. The latent variables were z-standardized.

It is assumed that SFON and symbolic number skills measured at the beginning of Grade 1 is something that takes place at the within level. The between level does not seem to be of importance when answering the research questions because the children came into each Grade 1 class from multiple kindergarten groups, reducing the influence of class at the first measurement point. Therefore, a single-level model with cluster-robust standard errors was estimated. In addition, a single-level model was also used for empirical reasons such as a low interclass correlation (ICC = 0.025–0.045) for the SFON indicators. Model fit was evaluated using multiple fit indices. CFI values >0.95, RMSEA values <0.06, SRMR values <0.08 (Weiber and Mühlhaus, 2014) and $\chi^2/df < 3$ (Homburg and Giering, 1996) indicate a good model fit.

RESULTS

Descriptive Statistics and Correlations

Table 2 presents the descriptive statistics for all measures which highlights the large variance between the children.

In **Figure 3**, the results demonstrate that SFON is heterogeneous. 26.4% of the children never gave a numerical answer. Only 1% of the children achieved the maximum SFON score. Also, the mean of 3.38 ($SD = 3.42$) indicates that children seem to pay little attention to the numerical aspects of the pictures at the beginning of Grade 1. To test whether the children gave more numerical answers to any one of the three SFON pictures, an ANOVA with repeated measures was conducted. The result showed significant differences in the mean values of the three SFON pictures [$F(1.91, 2{,}395.40) = 245.00$, $p < 0.001$, $\eta_p^2 = 0.16$]. A post-hoc test showed significant differences between all items ($p < 0.001$), with an increase of SFON-answers from picture 1 to picture 3.

Correlation analyses (**Table 3**) indicate a significant, but very weak association ($r < 0.2$) between SFON tendency and symbolic number skills ($r = 0.18$, $p = 0.000$), mathematical achievement at the end of Grade 1 ($r = 0.17$, $p = 0.000$), working memory ($r = 0.11$, $p = 0.000$), nonverbal IQ ($r = 0.15$, $p = 0.000$) and first language ($r = -0.07$, $p = 0.024$). The strongest correlation was found between symbolic number skills and the mathematical achievement at the end of Grade 1 ($r = 0.55$, $p = 0.000$). Symbolic number skills are moderately correlated with working memory ($r = 0.48$, $p = 0.000$) and nonverbal IQ ($r = 0.44$, $p = 0.000$). Mathematical achievement at the end of Grade 1 was also significantly correlated with working memory ($r = 0.50$, $p = 0.000$) and nonverbal IQ ($r = 0.48$, $p = 0.000$). The correlation between working memory and nonverbal IQ was moderate with $r = 0.45$ ($p = 0.000$) (Cohen, 1992). All other correlations were weak.

TABLE 3 | Correlations between all variables.

	1	*2*	*3*	*4*	*5*	*6*
1. SFON						
2. Symbolic number skills	0.18***					
3. Mathematical achievement	0.17***	0.55***				
4. Working memory	0.11***	0.48***	0.50***			
5. Nonverbal IQ	0.15***	0.44***	0.48***	0.45***		
6. First language	−0.07*	−0.12***	−0.15***	−0.09**	−0.07*	
7. Gender	0.03	0.02	0.15***	−0.03	−0.01	0.03

Note. *p < 0.05; **p < 0.01; ***p < 0.001.

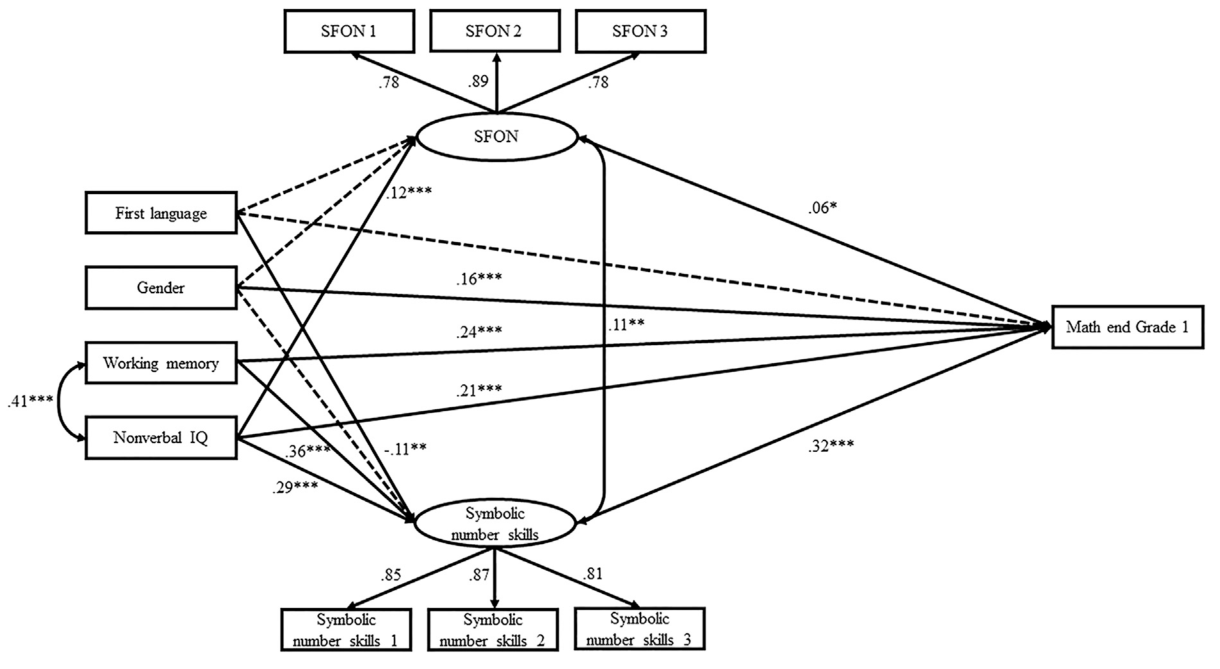


FIGURE 4 | Structural equation model of the final model, containing all hypothesized paths and covariances. Solid arrows represent the hypothesized significant paths. Dashed arrows depict paths that were not significant. Standardized estimates are provided with their levels of significance. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Structural Equation Model

The model with SFON and symbolic number skills as latent variables fitted the data well, $\chi^2(34) = 68.37$, $p < 0.05$, CFI = 0.99, RMSEA = 0.03, SRMR = 0.03. The estimated parameters are presented in **Figure 4**. SFON and symbolic number skills were significant predictors of mathematical achievement at the end of Grade 1 when first language, gender, working memory and nonverbal IQ were controlled for. The effect of symbolic number skills ($\beta = 0.32$, $p = 0.000$) was higher than the effect of SFON ($\beta = 0.06$, $p = 0.020$). The correlation between SFON and symbolic number skills was weak ($r = 0.11$, $p = 0.001$). First language ($\beta = -0.06$, $p = 0.130$) and gender ($\beta = 0.04$, $p = 0.233$) did not predict SFON. But nonverbal IQ did have an effect on SFON ($\beta = 0.12$, $p = 0.000$). Furthermore, gender did not predict symbolic number skills. First language, working memory and nonverbal IQ were significant predictors of symbolic number skills. Children who had German as a first language had a significantly higher score for symbolic number skills. In addition, gender, working memory and nonverbal IQ predicted mathematical achievement. Boys reached higher mathematical achievement than girls at the end of Grade 1. Working memory was correlated with nonverbal IQ ($r = 0.41$, $p = 0.000$).

A comparative model was constructed to test whether the effect of SFON on mathematical achievement at the end of Grade 1 differs significantly from that of symbolic number skills. In the comparative model, the paths between SFON and mathematical achievement, and symbolic number skills and mathematical achievement, were constrained to be equal and compared with the original model. In order to assess significant model differences, the two chi-square values were compared. The difference of $\Delta\chi^2 = 27.67$ ($\Delta df = 1$, $p = 0.000$) (Urban and Mayerl, 2014) suggests that the original model leads to an improvement of the model fit. The effect of symbolic number skills on mathematical achievement is therefore higher than the effect of SFON.

To test the stability of the model with SFON and symbolic number skills as latent variables, an alternative model was run with SFON and symbolic number skills as manifest variables. A single-level model with cluster-robust standard errors was evaluated. The modified model also fitted the data well, $\chi^2(6) = 15.35$, $p < 0.05$, CFI = 0.99, RMSEA = 0.04, SRMR = 0.03. The effects were identical to those of the original model.

DISCUSSION

This study investigated how a child's SFON and symbolic number skills, measured at the beginning of Grade 1, might predict their mathematical achievement at the end of Grade 1, controlling for nonverbal IQ, working memory, gender and first language. This relationship between SFON and symbolic number skills is interesting because while SFON focuses on the spontaneous recognition of small numbers, symbolic numerical knowledge (number words, exact numeration) can be improved through education.

The study also examined the extent to which children spontaneously focus on numerosity at the beginning of Grade 1.

Previous research had demonstrated that SFON is related to early numerical skills and subsequent mathematical achievement (e.g., Hannula and Lehtinen, 2005; Hannula-Sormunen et al., 2015; Nanu et al., 2018). But symbolic number skills, including skills such as counting, object counting, linking small magnitudes and numbers, are also significant predictors of mathematical achievement (e.g., Conoyer et al., 2016; Gallit et al., 2018). Symbolic skills have been shown to be crucial for later mathematical skills (e.g., Kolkman et al., 2013; Göbel et al., 2014).

In our study, we found that the SFON scores of the sample were relatively low at the beginning of Grade 1. About a quarter of the children did not focus spontaneously on the numerical aspects of the pictures and only a few achieved the maximum SFON score.

Correlation analyses indicated a significant, but weak relationship between SFON and symbolic number skills ($r = 0.18$, $p = 0.000$). The correlation based on the structural equation model was $r = 0.11$ ($p = 0.001$). These correlations are lower than in other studies (e.g., Hannula and Lehtinen, 2005; Edens and Potter, 2013). Hannula and Lehtinen (2005) identified a significant correlation between children's SFON and their number sequence skills ($r = 0.42$, $p < 0.01$) and object counting skills ($r = 0.35$, $p < 0.01$). Edens and Potter (2013), found that the correlation between SFON and counting skills was $r = 0.71$ ($p < 0.01$).

A couple of key reasons could account for the lower SFON scores and poor correlations found by us compared to the results reported by other SFON studies. In this study the assessment took place at the beginning of Grade 1, while in most other studies SFON was measured earlier, before school entry. Hannula and Lehtinen (2005) showed that SFON is a stable construct, so reasons other than age could account for the low SFON scores. It is likely that structured numerical instruction, which begins at age 4 in Switzerland with a compulsory kindergarten curriculum, may have influenced the result. Children begin first grade with a rather high level of symbolic skills so spontaneous focusing on small sets of items might be less important. Also, most of the studies that reported higher SFON scores used action-based tasks (e.g., Hannula and Lehtinen, 2005; Hannula et al., 2010; Nanu et al., 2018). Therefore, the outcome could have been influenced by the selection of a verbally-based picture task and its concomitant limitations. For future studies, it would be important to use both action- and verbally-based tasks and examine whether a different format of task results in different outcomes.

Results based on the structural equation model showed that both SFON and symbolic number skills significantly predicted mathematical achievement at the end of Grade 1. But the effect of number skills on mathematical achievement at the end of Grade 1 was much higher ($\beta = 0.32$, $p = 0.000$) than the effect of SFON ($\beta = 0.06$, $p = 0.020$). As in many other studies (e.g., Missall et al., 2012; Kolkman et al., 2013; Göbel et al., 2014; Toll et al., 2016; Caviola et al., 2020), the findings support the hypothesis that symbolic number skills are a very important predictor for subsequent mathematical achievement.

The finding of a small effect of SFON on mathematical achievement at the end of Grade 1 agrees with the findings of Hannula et al. (2010), but not with the findings of Hannula-Sormunen et al. (2015). This difference might be because the later study assessed fewer control variables. Other reasons that may explain the differences between the results of the Hannula-Sormunen et al. (2015) and the present study are: First, different tasks were used to measure SFON (verbally-vs. action-based). Second, in the Hannula-Sormunen et al. (2015) study only counting skills and subitizing were assessed to determine number skills, whereas in the present study, symbolic number skills were assessed using multiple tasks (e.g. counting objects and linking the result with the correct number, comparing numbers up to 20, number sequences up to 20, addition and subtraction). Third, the sample size in the Hannula-Sormunen et al. (2015) study was very small ($N = 36$). And finally, working memory, which might be crucial when carrying out an action-based SFON task, was not assessed.

The influence of the control variables on SFON and symbolic number skills confirms findings reported by other researchers. Nonverbal IQ and working memory affect symbolic number skills and mathematical achievement at the end of Grade 1 (e.g., De Smedt et al., 2009; Krajewski and Schneider, 2009). IQ also predicts SFON. Boys outperformed girls in mathematical achievement at the end of Grade 1, but not in SFON and symbolic number skills measured at the beginning of Grade 1. This corresponds with the findings of other studies (e.g., Niklas and Schneider, 2012; Sale et al., 2018) that the relationship between gender and mathematical performance in young children remains unclear. In addition, the picture task requires language competence therefore, first language was included as a control variable. But no effect of the children's first language on SFON was found and it can be assumed that language competence did not affect the result. However, it would be important, to assess language competence with more differentiated measures, such as vocabulary or the knowledge of number words in future studies. The influence of first language on other numerical and mathematical constructs is harder to unpick. Number skills were predicted by first language, and bilingual children and children with German as a first language had higher symbolic number skills scores. But this was not the case for mathematical achievement at the end of Grade 1. Language requirements were low in this test, and it may be that the children with a first language other than German improved their language skills during the first year of school.

The study had some limitations. First, SFON was measured using the picture task, which is a verbally-based task. According to Batchelor et al. (2015), SFON scores are affected by the type of task used during assessments. Therefore, it is possible that an action-based task, like a selection task, could have produced different results. As Hannula and Lehtinen (2005) emphasize, it is important to use varied SFON measures to get a reliable indicator of children's SFON. In addition, symbolic number skills were measured with a battery of sub-tests with multiple items. Therefore, to compare the effect of SFON on mathematical achievement gain, it would have been useful to also measure SFON with multiple items. Unfortunately, due to time and funding constraints, this was not possible in this study. Second, this type of SFON task requires language skills and a knowledge of number words and other vocabulary, which could influence the responses, especially those of second language learners or children with a language impairment. To deal with this problem, language competence was included in the model. The study was unable to use a more sensitive measure of language competence because of constraints. Third, the test instruction "This is the first picture" includes numerical information, which might have steered children to focus on the numerical aspects of the picture. However, the rate of numerical answers provided to the first item was very low, and it seems that the hint did not affect the children's answers. Fourth, the third picture shows a typical numerical board game situation, which again might have caused the children to focus on numbers. This picture did have the highest number of numerical responses. There was, however, also an increase in such statements between picture 1 and picture 2. Therefore, the increase could simply be due to a habituation effect. Nevertheless, the validity of the SFON measurement could be improved. Further studies with more, revised, SFON tasks are necessary. Fifth, SFON and symbolic number skills were measured at the beginning of Grade 1, a later point than that used by most other studies. It could be that this relationship would be different at the beginning of kindergarten. Finally, on a methodological level, homogenous parceling was not possible for the symbolic number skills construct.

Given these limitations, taking into account that SFON also requires symbolic numerical skills like exact numeration, more research is needed to disentangle the complex relationship between SFON and other mathematical skills and its impact on the mathematical learning process.

IMPLICATIONS AND FURTHER RESEARCH

To the best of our knowledge, this is the first longitudinal study to use a large sample of more than 1,000 first graders to investigate how the relationship between SFON and a broad range of symbolic number skills influences mathematical achievement gains. The study confirms previous research findings. Both SFON and symbolic number skills predict mathematical achievement at the end of Grade 1, although symbolic number skills have a much stronger effect. These results have implications for mathematical education in kindergarten. They highlight the importance of using measures to foster symbolic number skills in general and the important role of mathematical education programs. Structured programs (e.g., Krajewski et al., 2008; Ennemoser et al., 2015) and play-based interventions (e.g., Hauser et al., 2014; Jörns et al., 2014) have been proven to be successful. There is also evidence that children's SFON can be enhanced with guided interventions during everyday situations in day care settings too (Hannula et al., 2005; Braham et al., 2018; Hannula-Sormunen et al., 2020).

In future studies, it would also be interesting to examine if, and if so, how, the pre-school context influences how SFON and symbolic number effect mathematical achievement gain. For example, research by Kuratli Geeler (2019) has shown that children in Switzerland, which starts formal mathematical education in kindergarten, developed more symbolic numerical skills than children in kindergartens in Germany, where a child-oriented approach to early education dominates (Gasteiger et al., 2021). In addition, the present study has shown that more research into how SFON should be measured is required.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Zurich, Faculty of Arts and Social Sciences, Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

NG, DL and EMO conceptualized the research. NG and DL were responsible for the data collection. NG performed the statistical analyses, wrote the first draft, and finalized the manuscript. DL contributed substantially to the section on number skills. EMO supervised the analyses and helped to finalize the manuscript. All authors approved the submitted version.

FUNDING

Data collection was conducted within the project "Learning mathematics and student cooperation in grade 1 and 2", with financial support by the swiss universities "Development of Subject-Specific Education" programme (P-9).

REFERENCES

Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., et al. (2012). Home and Preschool Learning Environments and Their Relations to the Development of Early Numeracy Skills. *Early Child. Res. Q.* 27 (2), 231–244. doi:10.1016/j.ecresq.2011.08.003

Batchelor, S., Inglis, M., and Gilmore, C. (2015). Spontaneous Focusing on Numerosity and the Arithmetic Advantage. *Learn. Instruction* 40, 79–88. doi:10.1016/j.learninstruc.2015.09.005

Bojorque, G., Torbeyns, J., Hannula-Sormunen, M., Van Nijlen, D., and Verschaffel, L. (2017). Development of SFON in Ecuadorian Kindergartners. *Eur. J. Psychol. Educ.* 32 (3), 449–462. doi:10.1007/s10212-016-0306-9

Braham, E. J., Libertus, M. E., and McCrink, K. (2018). Children's Spontaneous Focus on Number before and after Guided Parent-Child Interactions in a Children's Museum. *Dev. Psychol.* 54 (8), 1492–1498. doi:10.1037/dev0000534

Caviola, S., Colling, L. J., Mammarella, I. C., and Szűcs, D. (2020). Predictors of Mathematics in Primary School: Magnitude Comparison, Verbal and Spatial Working Memory Measures. *Dev. Sci.* 23, e12957. doi:10.1111/desc.12957

Chan, J. Y., and Mazzocco, M. M. (2017). Competing Features Influence Children's Attention to Number. *J. Exp. Child. Psychol.* 156, 62–81. doi:10.1016/j.jecp.2016.11.008

Cohen, J. (1992). A Power Primer. *Psychol. Bull* 112 (1), 155–159. doi:10.1037//0033-2909.112.1.155

Conoyer, S. J., Foegen, A., and Lembke, E. S. (2016). Early Numeracy Indicators. *Remedial Spec. Educ.* 37 (3), 159–171. doi:10.1177/0741932515619758

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., and Ghesquière, P. (2009). Working Memory and Individual Differences in Mathematics Achievement: A Longitudinal Study from First Grade to Second Grade. *J. Exp. Child. Psychol.* 103 (2), 186–201. doi:10.1016/j.jecp.2009.01.004

Dehaene, S. (2001). Precis of the Number Sense. *Mind Lang.* 16 (1), 16–36. doi:10.1111/1468-0017.00154

Deutschschweizer Erziehungsdirektoren-Konferenz (D-EDK) (2016). Lehrplan 21. Mathematik. Kompetenzaufbau 1. Zyklus [Curriculum 21. Mathematics. Development of competence. First cycle]. Available at: https://v-fe.lehrplan.ch/lehrplan_printout.php?k=1&z=1&ekalias=0&fb_id=5 (Accessed November 12, 2020).

Dornheim, D. (2008). Prädiktion von Rechenleistung und Rechenschwäche: der Beitrag von Zahlen-Vorwissen und allgemein-kognitiven Fähigkeiten *[Prediction of arithmetical achievement and learning difficulties in mathematics: The contribution of numerical knowledge and cognitive ability]*. Berlin: Logos-Verlag.

Dowker, A. (2005). *Individual Differences in Arithmetic: Implications for Psychology, Neuroscience, and Education*. New York: Psychology Press.

Edens, K. M., and Potter, E. F. (2013). An Exploratory Look at the Relationships Among Math Skills, Motivational Factors and Activity Choice. *Early Child. Educ J* 41 (3), 235–243. doi:10.1007/s10643-012-0540-y

Ennemoser, M., Sinner, D., and Krajewski, K. (2015). Kurz- und langfristige Effekte einer entwicklungsorientierten Mathematikförderung bei Erstklässlern mit drohender Rechenschwäche. *Lernen und Lernstörungen* 4 (1), 43–59. doi:10.1024/2235-0977/a000091

Frydman, O. (1995). The Concept of Number and the Acquisition of Number Concepts: The 'when', the 'how', and the 'what' of it. *Cah. Psychol. Cogn.* 14 (6), 653–684.

Fuson, K. (1988). *Children's Counting and Number Concept*. New York: Springer.

Gallit, F., Wyschkon, A., Poltz, N., Moraske, S., Kucian, K., von Aster, M., et al. (2018). Henne oder Ei: Reziprozität mathematischer Vorläufer und Vorhersage des Rechnens [Chicken or egg: Reciprocity of mathematical precursor skills and the prediction of mathematics school achievement]. *Lernen und Lernstörungen* 7 (2), 81–92. doi:10.1024/2235-0977/a000205

Gasteiger, H., Brunner, E., and Chen, C.-S. (2021). Basic Conditions of Early Mathematics Education-A Comparison between Germany, Taiwan and Switzerland. *Int. J. Sci. Math. Educ.* 19, 111–127. doi:10.1007/s10763-019-10044-x

Göbel, S. M., Watson, S. E., Lervåg, A., and Hulme, C. (2014). Children's Arithmetic Development: It Is Number Knowledge, Not the Approximate Number Sense, that Counts. *Psychol. Sci.* 25 (3), 789–798. doi:10.1177/0956797613516471

Gray, S. A., and Reeve, R. A. (2016). Number-Specific and General Cognitive Markers of Preschoolers' Math Ability Profiles. *J. Exp. Child. Psychol.* 147, 1–21. doi:10.1016/j.jecp.2016.02.004

Hannula, M. M., Lepola, J., and Lehtinen, E. (2010). Spontaneous Focusing on Numerosity as a Domain-Specific Predictor of Arithmetical Skills. *J. Exp. Child. Psychol.* 107 (4), 394–406. doi:10.1016/j.jecp.2010.06.004

Hannula, M. M., and Lehtinen, E. (2005). Spontaneous Focusing on Numerosity and Mathematical Skills of Young Children. *Learn. Instruction* 15 (3), 237–256. doi:10.1016/j.learninstruc.2005.04.005

Hannula, M. M., Mattinen, A., and Lehtinen, E. (2005). ""Does Social Interaction Influence 3-Year-Old Children's Tendency to Focus on Numerosity? A Quasi-Experimental Study in Day Care," in *Powerful Learning Environments for Promoting Deep Conceptual and Strategic Learning*. Editors L. Verschaffel, E. De Corte, G. Kanselaar, and M. Valcke (Leuven: Leuven University Press), 63–80.

Hannula, M. M., Räsänen, P., and Lehtinen, E. (2007). Development of Counting Skills: Role of Spontaneous Focusing on Numerosity and Subitizing-Based Enumeration. *Math. Thinking Learn.* 9 (1), 51–57. doi:10.1207/s15327833mtl0901_4

Hannula, M. M. (2005). Spontaneous Focusing on Numerosity in the Development of Early Mathematical Skills. [dissertation thesis]. Finland, Turku: University of Turku. Available at: https://www.utupub.fi/handle/10024/102190 (Accessed November 12, 2020).

Hannula-Sormunen, M. M., Lehtinen, E., and Räsänen, P. (2015). Preschool Children's Spontaneous Focusing on Numerosity, Subitizing, and Counting Skills as Predictors of Their Mathematical Performance Seven Years Later at School. *Math. Thinking Learn.* 17 (2–3), 155–177. doi:10.1080/10986065.2015.1016814

Hannula-Sormunen, M., Nanu, C., Luomaniemi, K., Heinonen, M., Sorariutta, A., Södervik, I., et al. (2020). Promoting Spontaneous Focusing on Numerosity and Cardinality-Related Skills at Day Care with One, Two, How many and Count, How many Programs. *Math. Thinking Learn.* 22 (4), 312–331. doi:10.1080/10986065.2020.1818470

Hasselhorn, M., Schumann-Hengsteler, R., Gronauer, J., Grube, D., Mähler, C., Schmid, et al. (2012). AGTB 5-12. Arbeitsgedächtnistestbatterie für Kinder von 5 bis 12 Jahren *[AGTB 5-12. Working memory test battery for children from the age of 5 to 12 years]*. Göttingen: Hogrefe.

Hauser, B., Vogt, F., Stebler, R., and Rechsteiner, K. (2014). Förderung Früher Mathematischer Kompetenzen. *Frühe Bildung* 3 (3), 139–145. doi:10.1026/2191-9186/a000144

Hornburg, C., and Giering, A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. *Marketing (Münch.)* 18 (1), 5–24. doi:10.15358/0344-1369-1996-1-5

Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting First-Grade Math Achievement from Developmental Number Sense Trajectories. *Learn. Disabil Res Pract* 22 (1), 36–46. doi:10.1111/j.1540-5826.2007.00229.x

Jörns, C., Schuchardt, K., Grube, D., and Mähler, C. (2014). Spielorientierte Förderung numerischer Kompetenzen im Vorschulalter und deren Eignung zur Prävention von Rechenschwierigkeiten [Playful training of numerical skills in kindergarten for prevention of mathematical disability]. *Empir. Sonderpädag.* 3, 243–259. doi:10.25656/01:9933

Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early Numerical Development and the Role of Non-symbolic and Symbolic Skills. *Learn. Instruction* 25, 95–103. doi:10.1016/j.learninstruc.2012.12.001

Krajewski, K., Nieding, G., and Schneider, W. (2008). Kurz- und langfristige Effekte mathematischer Frühförderung im Kindergarten durch das Programm "Mengen, zählen, Zahlen". *Z. für Entwicklungspsychologie Pädagogische Psychol.* 40, 135–146. doi:10.1026/0049-8637.40.3.135

Krajewski, K., and Schneider, W. (2009). Early Development of Quantity to Number-Word Linkage as a Precursor of Mathematical School Achievement and Mathematical Difficulties: Findings from a Four-Year Longitudinal Study. *Learn. Instruction* 19 (6), 513–526. doi:10.1016/j.learninstruc.2008.10.002

Krajewski, K. (2003). Vorhersage von Rechenschwäche in der Grundschule *[Disabilities in mathematics in primary school]*. Hamburg: Verlag Dr. Kovač.

Kucian, K., Kohn, J., Hannula-Sormunen, M. M., Richtmann, V., Grond, U., Käser, T., et al. (2012). Kinder mit Dyskalkulie fokussieren spontan weniger auf

Anzahligkeit [Children with developmental dyscalculia spontaneously focus less on numerosity]. *Lern. Lernstör.* 1 (4), 241–253. doi:10.1024/2235-0977/a000024

Kuratli Geeler, S. (2019). Mathematische Kompetenzen von Kindergartenkindern: Überprüfung eines Testinstrumentes und Analyse von Unterschieden in der numerischen Leistungsentwicklung [Mathematical skills of kindergarteners: Examining a test and analyzing the differences in numerical achievement gain of kindergarten children]. [dissertation thesis]. Zurich, Switzerland: University of Zurich. doi:10.5167/uzh-171088

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., and Von Davier, M. (2013). Why Item Parcels Are (Almost) Never Appropriate: Two Wrongs Do Not Make a Right-Ccamouflaging Misspecification with Item Parcels in CFA Models. *Psychol. Methods* 18 (3), 257–284. doi:10.1037/a0032773

Matsunaga, M. (2008). Item Parceling in Structural Equation Modeling: A Primer. *Commun. Methods Measures* 2 (4), 260–293. doi:10.1080/19312450802458935

McMullen, J., Hannula-Sormunen, M. M., and Lehtinen, E. (2015). Preschool Spontaneous Focusing on Numerosity Predicts Rational Number Conceptual Knowledge 6 Years Later. *ZDM Maths. Educ.* 47 (5), 813–824. doi:10.1007/s11858-015-0669-4

Missall, K. N., Mercer, S. H., Martínez, R. S., and Casebeer, D. (2012). Concurrent and Longitudinal Patterns and Trends in Performance on Early Numeracy Curriculum-Based Measures in Kindergarten through Third Grade. *Assess. Eff. Intervention* 37 (2), 95–106. doi:10.1177/1534508411430322

Nanu, C. E., McMullen, J., Munck, P., Group, P. S., and Hannula-Sormunen, M. M. (2018). Spontaneous Focusing on Numerosity in Preschool as a Predictor of Mathematical Skills and Knowledge in the Fifth Grade. *J. Exp. Child. Psychol.* 169, 42–58. doi:10.1016/j.jecp.2017.12.011

Nanu, C., Laakkonen, E., and Hannula-Sormunen, M. (2020). The Effect of First School Years on Mathematical Skill Profiles. *FLR* 8 (1), 56–75. doi:10.14786/flr.v8i1.485

Niklas, F., and Schneider, W. (2012). Die Anfänge geschlechtsspezifischer Leistungsunterschiede in mathematischen und schriftsprachlichen Kompetenzen [The beginning of gender-based performance differences in mathematics and linguistic competencies]. *Z. Entwicklungspsychol. Pädagog. Psychol.* 44 (3), 123–138. doi:10.1026/0049-8637/a000064

Rathé, S., Torbeyns, J., De Smedt, B., and Verschaffel, L. (2019). Spontaneous Focusing on Arabic Number Symbols and its Association with Early Mathematical Competencies. *Early Child. Res. Q.* 48 (3), 111–121. doi:10.1016/j.ecresq.2019.01.011

Rathé, S., Torbeyns, J., Hannula-Sormunen, M. M., and Verschaffel, L. (2016). Kindergartners' Spontaneous Focusing on Numerosity in Relation to Their Number-Related Utterances during Numerical Picture Book reading. *Math. Think. Learn.* 18 (2), 125–141. doi:10.1080/10986065.2016.1148531

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* 48 (2), 1–36. doi:10.18637/jss.v048.i02

Sale, A., Schell, A., Koglin, U., and Hillenbrand, C. (2018). Einflussfaktoren mathematischer Kompetenzen vor Schuleintritt [Factors influencing mathematical competences before school entry]. *Empir. Sonderpädag.* 4, 370–387. doi:10.25656/01:16781

Simon, T. J., and Vaishnavi, S. (1996). Subitizing and Counting Depend on Different Attentional Mechanisms: Evidence from Visual Enumeration in Afterimages. *Percept. Psychophys.* 58 (6), 915–926. doi:10.3758/BF03205493

Szkudlarek, E., and Brannon, E. M. (2017). Does the Approximate Number System Serve as a Foundation for Symbolic Mathematics? *Lang. Learn. Dev.* 13 (2), 171–190. doi:10.1080/15475441.2016.1263573

Toll, S. W. M., Kroesbergen, E. H., and van Luit, J. E. H. (2016). Visual Working Memory and Number Sense: Testing the Double Deficit Hypothesis in Mathematics. *Br. J. Educ. Psychol.* 86 (3), 429–445. doi:10.1111/bjep.12116

Urban, D., and Mayerl, J. (2014). Strukturgleichungsmodellierung. Ein Ratgeber für die Praxis *[Structural equation modeling. A practical guide]*. Wiesbaden: Springer.

Weiber, R., and Mühlhaus, D. (2014). Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS *[Structural equation modeling. An application-oriented introduction to causal analyses using AMOS, SmartPLS and SPSS]*. Berlin: Springer Gabler.

Weiß, R. H., and Osterland, J. (2013). CFT 1-R: Grundintelligenztest Skala 1 *[CFT 1-R: Culture Fair* Intelligence *Tests - Scale 1]*. Göttingen: Hogrefe.

Wilson, M. (2005). *Constructing Measures. An Item Response Modeling Approach*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Wynn, K. (1995). Origins of Numerical Knowledge. *Math. Cognit.* 1, 35–60.

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fpsyg.2021.765709

Check for updates

Cross-Format Integration of Auditory Number Words and Visual-Arabic Digits: An ERP Study

Sabrina Finke[1], Ferenc Kemény[1,2], Francina J. Clayton[3], Chiara Banfi[1,4], Anna F. Steiner[1,5], Corinna M. Perchtold-Stefan[1], Ilona Papousek[1], Silke M. Göbel[3,6] and Karin Landerl[1,7,8]*

[1]Institute of Psychology, University of Graz, Graz, Austria, [2]Institute of Education and Psychology at Szombathely, Eötvös Loránd University, Budapest, Hungary, [3]Department of Psychology, University of York, York, United Kingdom, [4]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria, [5]FH JOANNEUM, University of Applied Sciences, Graz, Austria, [6]Department of Special Needs Education, University of Oslo, Oslo, Norway, [7]BioTechMed-Graz, Graz, Austria, [8]Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia

Converting visual-Arabic digits to auditory number words and vice versa is seemingly effortless for adults. However, it is still unclear whether this process takes place automatically and whether accessing the underlying magnitude representation is necessary during this process. In two event-related potential (ERP) experiments, adults were presented with identical (e.g., "one" and 1) or non-identical (e.g., "one" and 9) number pairs, either unimodally (two visual-Arabic digits) or cross-format (an auditory number word and a visual-Arabic digit). In Experiment 1 (N = 17), active task demands required numerical judgments, whereas this was not the case in Experiment 2 (N = 19). We found pronounced early ERP markers of numerical identity unimodally in both experiments. In the cross-format conditions, however, we only observed late neural correlates of identity and only if the task required semantic number processing (Experiment 1). These findings suggest that unimodal pairs of digits are automatically integrated, whereas cross-format integration of numerical information occurs more slowly and involves semantic access.

Keywords: ERP, numerical cognition, cross-format integration, symbolic numbers, N1, N400

OPEN ACCESS

Edited by:
Ricardo Moura, University of Brasilia, Brazil

Reviewed by:
Luca Rinaldi, University of Milano-Bicocca, Italy
Robert Reeve, The University of Melbourne, Australia

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

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

Received: *27 August 2021*
Accepted: *27 October 2021*
Published: *23 November 2021*

Citation:
Finke S, Kemény F, Clayton FJ, Banfi C, Steiner AF, Perchtold-Stefan CM, Papousek I, Göbel SM and Landerl K (2021) Cross-Format Integration of Auditory Number Words and Visual-Arabic Digits: An ERP Study. Front. Psychol. 12:765709. doi: 10.3389/fpsyg.2021.765709

INTRODUCTION

Whether number words and visual-Arabic digits are automatically and involuntarily linked is an enduring open question in cognitive psychology. Every day, we navigate through our modern literate world by integrating numerical information from different sensory modalities: Following the station announcement that our train leaves from platform five, we search for the corresponding visual-Arabic digit 5. In adults, switching from one number format to another seems to happen effortlessly. While this daily experience suggests an automatic integration of number words and digits, empirical evidence is still critically lacking. Therefore, the current paper addresses the question whether the mental representation of a specific digit is automatically and involuntarily activated upon hearing the corresponding number word. Here, we attempt to unravel the neurocognitive mechanisms underlying this integration process.

Already at a young age children acquire the skill to link verbal numbers to their written digit counterparts: Words for small magnitudes (e.g., "two dogs," "three little pigs") are among the first words children learn (Durkin et al., 1986), and there is evidence suggesting that some children understand the meaning of single visual-Arabic digits as early as 18 months (Mix, 2009). When children enter school, their production and comprehension of single visual-Arabic digits are already near perfect (Moura et al., 2015). Following many years of repeated exposure to numbers and frequent experience translating between different number formats, it seems intuitive that visual and verbal representations become increasingly linked. Representations of number words and digits may overlap to such an extent that one representation automatically activates the other. The strength of this link between number words and digits might be related to arithmetic performance in different age groups (children: Göbel et al., 2014; Malone et al., 2020; adults: Sasanguie and Reynvoet, 2014 but see Lyons et al., 2014; Sasanguie et al., 2017; Lin and Göbel, 2019). It has even been put forward that the mapping between number words and Arabic numbers might act as a "gatekeeper (or barrier) in the development of formal mathematical knowledge" (Purpura et al., 2013, p. 460).

The existence of a direct and automatic association of number words and digits was already proposed in the earliest version of the Triple Code Model (Dehaene and Cohen, 1995). According to this model, numbers are processed in three different numerical codes: The visual-Arabic number form processes numbers represented as digits, while spoken or written number words are represented in the verbal word frame. According to the model, both visual-Arabic digits and number words are symbols that do not *per se* contain any semantic information. Number semantics are only represented by the analogue magnitude representation, which is involved in all processes accessing the non-symbolic quantity of a given number. Bidirectional translational paths are postulated to directly link the different numerical codes. Crucially, the Triple Code Model includes an asemantic transcoding route between representations of visual-Arabic digits and number words that has been shown to rely on left hemispheric pathways (Dehaene and Cohen, 1995). In other words, there is evidence for a direct route between the symbolic numerical representations of number words and of visual-Arabic digits, without the use of an indirect route through the activation of the underlying magnitude representation. In contrast, semantic models propose that number words and digits are only indirectly linked *via* their underlying meaning in terms of numerical magnitude. Specifically, semantic models of transcoding (e.g., Power and Dal Martello, 1990; McCloskey, 1992) assume that the source number is first transformed into an abstract analogue magnitude, which in turn then is transformed into the target number.

However, we frequently employ symbolic numbers in the absence of any actual numerical meaning. The magnitudes underlying certain combinations of digits, for example post codes, PIN codes, and telephone country codes do not necessarily carry relevant magnitude information.[1] The view that number words and digits are linked directly without an intermediary step of access to number semantics is supported by asemantic transcoding models (e.g., Power and Dal Martello, 1997; Barrouillet et al., 2004; Dotan and Friedmann, 2018). For example, the ADAPT model (Barrouillet et al., 2004) suggests that a verbal number word is parsed until a single word unit is identified. Each word unit or chunk either corresponds to lexicalized or non-lexicalized elements. Lexicalized elements can be directly retrieved from long-term memory and consist of lexical primitives including single-digit numbers one to nine, as well as teens, decades, and separators, such as hundred and thousand. On the other hand, non-lexicalized elements (e.g., "238") require complementary procedures. These can be best described as an algorithmic transcoding strategy which serves as a back-up if direct memory retrieval fails. Critically though, neither lexicalized nor non-lexicalized elements require any access to the underlying number semantics during the entire process of linking number words and their corresponding visual-Arabic digits.

Researchers have argued for the existence of a direct link between number words and digits based on behavioral findings from number comparison and matching tasks (Lyons et al., 2012; Marinova et al., 2018). Lyons et al. (2012) instructed participants to indicate the larger of a pair of quantities, presented in a single format (visual-Arabic digits, written number words or dot arrays), in a mixed symbolic format (visual-Arabic digits and written number words), or in a mixed symbolic-non-symbolic format (visual-Arabic digits and dot arrays). Participants showed switch costs in terms of significantly longer response times for mixed non-symbolic-symbolic pairs compared to single-format pairs. However, no switch costs were observed when comparing mixed symbolic pairs with single-format symbolic pairs, suggesting that number words and digits are closely linked. Marinova et al. (2018) were able to extend these findings using a task that did not require explicit magnitude judgments: They showed that also in a number matching task in which participants had to judge whether two quantities were numerically identical, participants were slower to compare mixed non-symbolic and symbolic pairs (tone sequences and digits) than their mixed symbolic counterparts (auditory number words and visual digits).

The fact that the co-activation of purely symbolic representations was faster than the co-activation of symbolic and non-symbolic representations points to a direct link between the visual-Arabic number form and the verbal word frame. However, task demands in number matching tasks may also elicit an activation of semantic content, that is, the numerical value of abstract number symbols. Therefore, findings from number matching tasks only provide indirect and incomplete evidence for the direct link between number words and digits.

Neuroscientific studies offer another window into investigating the association between number words and digits. More specifically, using event-related potential (ERP) methodology allows us to examine the time course of cross-format integration. Although ERP evidence about the direct link between number words and digits is still critically lacking, previous studies identified the importance of the N1 and N400 components in the processing of numerical stimuli.

The parietal N1 component was reported to be sensitive to numerical distance (Temple and Posner, 1998) and numerical

[1]While the first number of the country codes reflect continental information (e.g., Zone 1: North and Central America, Zones 3–4: Europe), the full country codes do not represent topographical organization. Some neighboring countries, like Sweden and Norway, do have neighboring codes, but others do not.

identity (Liu et al., 2018). Specifically, small numerical distances elicited larger N1 amplitudes than larger numerical distances in number comparison tasks with dots and digits in both adults and children between 5 and 9 years (Temple and Posner, 1998). However, the finding of an N1 amplitude modulation by numerical distance was not consistent across studies, as others could not replicate this finding, neither for non-symbolic nor symbolic numbers (Libertus et al., 2007; Hyde and Spelke, 2009). Generally, an N1 component can provide evidence for automatic and asemantic processing.

Semantic processing of numerical information is thought to be reflected by the N400 ERP component (Niedeggen et al., 1999; Galfano et al., 2004; Szücs and Csépe, 2005; Paulsen and Neville, 2008; Szücs and Soltész, 2010; Pinhas et al., 2014). The N400 is a central negative component peaking at around 400 ms, generally known to be sensitive to semantic mismatch or unexpectedness (Kutas and Federmeier, 2011). In the domain of numerical processing, the N400 was found to be sensitive to numerical identity using non-symbolic paradigms (Paulsen and Neville, 2008)[2] and also cross-format number pairs: In a passive paradigm not requiring numerical judgments, children showed more negative N400 components to mismatch between visually presented analogous magnitudes and auditorily presented number words already at the age of 3 to 5 years (Pinhas et al., 2014). Interestingly though, the authors reported that this effect of numerical identity was only observable for children who were already able to count. Overall, this suggests that once children have acquired basic counting knowledge, they automatically activate information about non-symbolic quantities and verbal number words *via* the underlying number semantic, even if they are not actively required to do so.

It is important to note that the numerical N400 effect can be dissociated from earlier N2b effects of perceptual non-match: Increased negative amplitudes have been reported at central electrode sites around 300–400 ms for incorrect versus correct calculations in arithmetic verification (Niedeggen et al., 1999; Szücs and Soltész, 2010) as well as in implicit probe tasks (Galfano et al., 2009). In terms of polarity and topography, this numerical N400 effect is highly similar to the classical N400 effect, often considered to reflect Lexico-semantic processing (Szücs et al., 2007; Szücs and Soltész, 2010). In terms of timing, however, the peak of the numerical N400 typically occurs around 100 ms earlier than for linguistic stimuli (Bassok et al., 2009; Guthormsen et al., 2017).

In the present study, these well-documented ERP components of numerical processing were used to investigate the temporal characteristics of unimodal and cross-format processing of symbolic number representations. In particular, we set out to test whether visual-Arabic digits and auditory number words are directly linked without explicit magnitude judgments being required as proposed by the Triple Code Model (Dehaene and Cohen, 1995) and asemantic transcoding models (Power and Dal Martello, 1997; Barrouillet et al., 2004; Dotan and Friedmann, 2018). We employed an ERP paradigm to investigate the possibly automatic link between number words and digits, both with a unimodal visual and a cross-format auditory–visual condition. If a direct link between number words and digits exists, we expect ERP effects of numerical identity to be present in both unimodal and cross-format conditions. Considering the previous literature (e.g., Liu et al., 2018), we hypothesized to find a larger N1 component for numerically identical trials than for numerically non-identical trials at parietal electrode sites. We conducted two experiments varying the access to the underlying number semantics: Experiment 1 involved numerical decisions and thus required participants to access the underlying number semantics in order to solve the active task. Experiment 2 did not involve any numerical judgments, and thus, no semantic access was needed. Due to the involvement of number semantics, we predicted an N400 effect of numerical identity with more negative deflections for non-identical than identical number pairs for Experiment 1. Observing a similar N400 effect also in Experiment 2 would suggest that pairs of numbers are linked semantically even when no access to the number semantic is required. In summary, we investigated ERP effects in response to the integration of auditory number words and visual-Arabic digits. We contrasted ERP effects of numerical identity robustly associated with automatic processing (N1 component) and semantic processing (N400 component).

EXPERIMENT 1

Participants

The sample comprised 17 healthy volunteers recruited at the University of Graz, Austria (age: $M = 22.6$ years, $SD = 2.4$; 8 males and 9 females). They were all native speakers of German and had normal or corrected-to-normal vision, as well as normal hearing status. Initially, three more participants took part but had to be excluded from data analysis because of technical issues with EEG recording or because of noisy data. Psychology students received course credit for participation. The study was conducted in accordance with the Declaration of Helsinki, and ethical clearance was obtained from the ethics committee of the University of Graz. Participants provided written informed consent prior to participation.

We conducted a *post hoc* sensitivity analysis with the "pwr" package (Champely, 2020) in R (R Core Team, 2020). Note that there were no estimates of effect sizes available in the literature, as the present research question had not been investigated previously. Our sensitivity analysis revealed that we would have been able to detect an effect of $\eta_p^2 = 0.15$ at an α of 0.05 and power set at 0.80 with the present sample size. As shown in the results section, the effects of numerical identity we observed for both components were even larger, which supports the adequacy of the current sample size.

Stimuli and Procedure

Participants were presented with pairs of numbers from 1 to 9 representing either the same or different numerosities. They were asked to indicate *via* keypress whether the second number of a pair was larger or smaller than five. The paradigm consisted

[2]This was also qualified by numerical distance, but as described below, numerical distance cannot be tested in our study.

of a unimodal and a cross-format block. In the unimodal block, the number pairs consisted of two visual-Arabic digits. In the cross-format block, the first number of a pair was a spoken number word, whereas the second number was a visual-Arabic digit. Each block was preceded by four practice trials to ensure that participants understood the task.

In each of the two blocks, participants were exposed to 240 number pairs appearing in a pseudorandom order. Number pairs contained digits and number words corresponding to the numerosities 1, 2, 4, 6, 8, and 9. As "seven" is a disyllabic number word in German, this numerosity was not included. To obtain an identical number of numerosities below and above five, we also decided not to include the numerosity 3. Visual stimuli, that is, Arabic digits, were presented in white on a black background with a height of 3 degrees of visual angle. In 120 trials per block, the number pairs were numerically identical, meaning that both numbers of a pair corresponded to the same numerosity. In the other 120 trials of a block, the number pairs were numerically non-identical. The numerical distance between the non-identical number pairs was either small (numerical distance of 1–3) or large (numerical distance of 5–8). Each block contained 60 number pairs with small and large numerical distances, respectively. In order to avoid low-level perceptual adaptation effects, we displayed visual-Arabic digits in one of four different spatial locations at 1 degree from the center of the display, using one of four different fonts (similar to an fMRI study by Vogel et al. (2017): Arial, Calibri, Century, Times New Roman). We ensured that in each trial, both constituents of a digit pair differed in terms of spatial locations and fonts. Number words were each presented by one of four speakers (two male and two female voices).

As shown in **Figure 1**, each trial began with a blank screen, displayed for 800 ms. Then, the first number was presented for 500 ms, before the second number appeared. This period was determined by the fact that this was the shortest possible period to present spoken number words comprehensively. Participants were asked to press the up arrow (right index finger) if the second number was larger than five or the down arrow (left index finger) if it was smaller than five. The next trial began as soon as a response was registered. If no response occurred within 2000 ms after stimulus onset, a question mark was displayed for another 3,000 ms. If participants did not respond within 5 s of stimulus onset, the next trial was presented.

ERP Recording and Data Analysis

Participants were seated in an acoustically and electrically shielded booth, 74 cm from the center of a 1920×1080 screen (refresh rate of 144 Hz). The paradigm was programmed with PsychoPy, version 1.90.1 (Peirce, 2009). Auditory stimuli were played by standard PC speakers. EEG was recorded from 19 Brain Products™ actiCAP active electrodes, positioned to the international 10–20 system using a BrainVision actiCHamp Research Amplifier (Brain Products™) with a sampling rate of 1,000 Hz and a stretchable electrode cap, referenced to the nose, and re-referenced offline to a mathematically averaged ears reference (Hagemann, 2004). We measured vertical and horizontal electrooculograms (EOGs) with two bipolar channels. Electrode impedances were below 30 kΩ for all electrodes. The continuous EEG was filtered (low cutoff: 0.1 Hz, time constant: 15.91, 24 dB/Oct; high cutoff: 100 Hz, 24 dB/Oct; notch filter: 50 Hz). EOG artifacts were removed by automatic ocular correction, using an ICA algorithm as implemented in BrainVision Analyzer 2.1 (slope mean, over the whole data, ICA with infomax algorithm, total squared correlations to delete: 30%; Gratton et al., 1983). Other artifacts were excluded automatically (gradient criteria: more than 50 μV difference between two successive data points or more than 200 μV difference in a 200 ms window; low activity criterion: less than 0.5 μV activity in a 100 ms window). The data were segmented into epochs of 700 ms before onset of the second number of a pair to the end of the trial (1,000 ms after stimulus onset). Because the first number of a pair was presented 500 ms before the second number, the time window of −700 to −500 ms served as the basis for baseline correction. Only segments with a correct response in a time window from 200 to 2000 ms were considered. All participants had at least 98 valid segments in each of the four conditions. On average, 112.82 ($SD = 4.22$) numerically identical segments and 113.29 ($SD = 4.84$) non-identical segments were retained for the unimodal conditions. For the cross-format conditions, an average of 114.94 ($SD = 4.98$) numerically identical and 113.53 ($SD = 5.48$) non-identical segments were included.

For the analysis of the N1 component, based on a previous ERP study of numerical identity (Liu et al., 2018), we averaged across a parietal electrode group including electrodes over left and right hemispheres (P3, P4, Pz, P7 and P8). The respective time window was identified as between 100 and 200 ms after stimulus onset. For the analysis of the N400 component, based on previous numerical processing ERP studies (Szücs and Csépe, 2005; Avancini et al., 2014), we considered a central electrode cluster (C3, C4, Cz). For the N400 component, we identified a time window from 250 to 400 ms after stimulus onset. For the N1 and N400 components, the peak was determined by detecting the most negative amplitude in the given time window for each electrode, and peak amplitude was defined as the average amplitude at peak and +/− 10 ms around the peak.

All statistical analyses were carried out with numerical identity (non-identical versus identical) and modality (unimodal versus cross-format) as within-subject variables. In all ANOVAs, we conducted separate follow-up analyses for each modality condition, even in the absence of a significant interaction to confirm that the main effects were not driven by only one of the modality conditions, but were reliable in both.

In principle, numerical distance effects can be used to test semantic access, more specifically by contrasting ERP effects for small and large numerical distances. However, this is not possible in the current design, because numerical distance was confounded with response selection: As the active task required participants to judge whether the second number of the pair was larger or smaller than five, number pairs with large numerical distance were always incongruent in terms of response selection. In other words, for number pairs with a large numerical distance, one number was always smaller than five, while the other was always larger than five. For number pairs with a

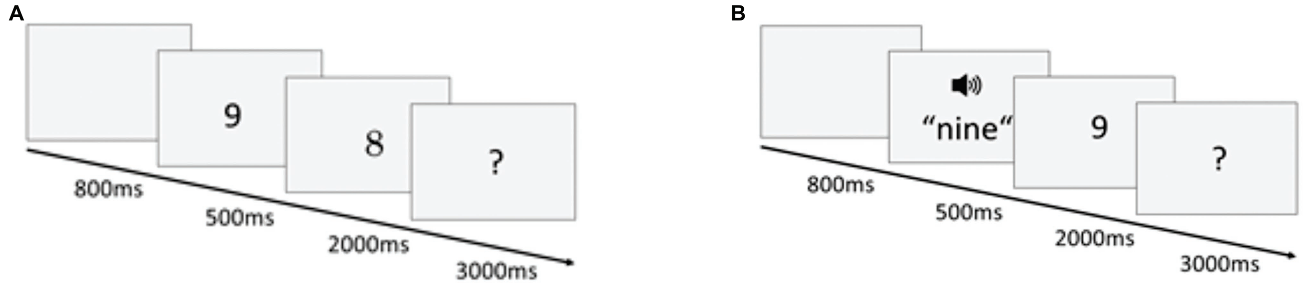


FIGURE 1 | Schematic time course of **(A)** a unimodal trial with a numerically non-identical visually presented number pair and **(B)** a cross-format trial with a numerically identical auditory–visual number pair.

small numerical distance, both numbers of a pair were often congruent in terms of response selection (both numbers either smaller or larger than five) – although this was not true for all cases (e.g., number pair "4" and "6"). Due to these confounds, we did not investigate numerical distance.

Results

The data collected for this study are publicly available on the Open Science Framework and can be accessed at https://osf.io/p7ksn/

Behavioral Measures: Accuracy and Reaction Times for the Numerical Decision Task

In a first step, we investigated participants' behavioral performance on our novel experimental task. As expected for this simple task format, accuracy was above 94% for both the unimodal and cross-modal conditions of the numerical decision task and was not further analyzed. Only RTs for correct responses were considered for analyses. We calculated median RTs for numerically identical as well as numerically non-identical number pairs separately for the unimodal and cross-format blocks for all participants. RTs by numerical identity and experimental block are provided in **Figure 2**.

We conducted a two-way repeated measures ANOVA with the within-subject factors identity (identical versus non-identical number pairs) and modality (unimodal versus cross-format). The ANOVA revealed a significant main effect of identity, $F(1,16) = 35.95$, $p < 0.001$, $\eta_p^2 = 0.69$, with higher RTs for non-identical versus identical number pairs. Neither the main effect of modality nor the identity x modality interaction were significant (both ps > 0.341). As described above, we ran separate repeated measures ANOVAs for each modality with identity as within-subject variable. A significant effect of identity was confirmed for both the unimodal, $F(1,17) = 14.06$, $p = 0.002$, $\eta_p^2 = 0.47$, and the cross-format condition, $F(1,17) = 21.68$, $p < 0.001$, $\eta_p^2 = 0.58$.

N1

The averaged waveforms of the parietal electrode cluster by identity and modality are depicted in **Figure 3**. An identity x modality ANOVA revealed a significant main effect of identity, $F(1,16) = 5.10$, $p = 0.038$, $\eta_p^2 = 0.24$, with more negative peak amplitudes for identical than non-identical number pairs. The interaction was also significant, $F(1,16) = 4.58$, $p = 0.048$, $\eta_p^2 = 0.22$. The main effect of modality was not significant ($F < 1$).

To further analyze the identity x modality interaction, a separate repeated measures ANOVA was conducted for each modality with identity (identical vs. non-identical) as within-subject variable. The ANOVAs revealed a significant effect for the unimodal block, $F(1,16) = 8.62$, $p = 0.010$, $\eta_p^2 = 0.35$, with more negative peak amplitudes for identical than non-identical number pairs. For the cross-format block, however, the difference between identical and non-identical pairs was not significant, $F(1,16) = 0.02$, $p = 0.879$, $\eta_p^2 = 0.00$.

N400

Figure 4 provides the averaged waveforms of the central electrode cluster by identity and modality. As can be seen, the amplitudes of the waveforms were more negative for non-identical than identical number pairs in both unimodal and cross-format blocks. A 2 × 2 repeated measures ANOVA was conducted with identity (identical vs. non-identical) and modality (unimodal vs. cross-format) as within-subject factors. There was a significant main effect of identity, $F(1,16) = 11.19$, $p = 0.004$, $\eta_p^2 = 0.41$, with more negative peak amplitudes for non-identical than identical number pairs. The main effect of modality was also significant, $F(1,16) = 11.51$, $p = 0.004$, $\eta_p^2 = 0.42$. For the cross-format block, the peak amplitudes were more negative than for the unimodal block. The interaction identity x modality was not significant ($F < 1$).

To ensure that the identity-based effect was present in both modality conditions, we conducted two separate ANOVAs with identity (identical vs. non-identical) as within-subject variable. For the unimodal block, there were more negative peak amplitudes for non-identical than identical number pairs, $F(1,16) = 7.11$, $p = 0.017$, $\eta_p^2 = 0.31$. Similarly, in the cross-format block, there were more negative peak amplitudes for non-identical than identical number pairs, $F(1,16) = 7.75$, $p = 0.013$, $\eta_p^2 = 0.33$.

Discussion

The behavioral and electrophysiological results of Experiment 1 support the view that number words and digits are linked *via* number semantics when explicit numerical decisions are required.

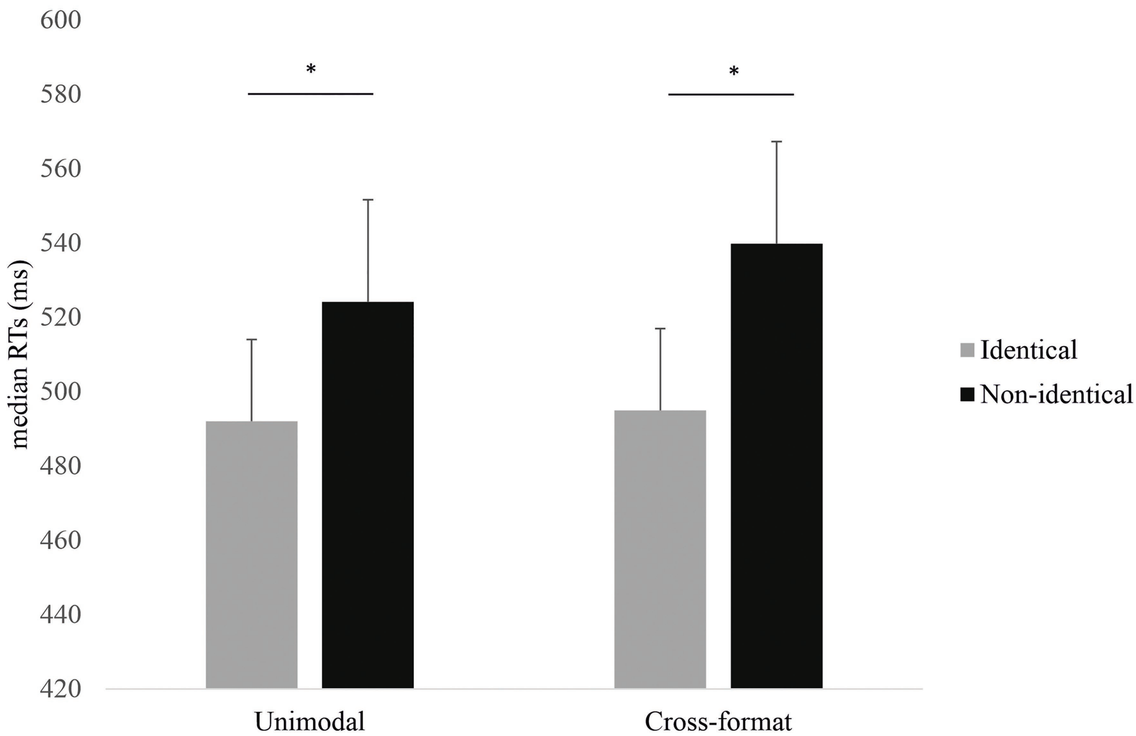


FIGURE 2 | Median reaction times (RTs) by numerical identity and modality. Error bars indicate SEM. Asterisks indicate significant differences ($p < 0.05$).

Behaviorally, participants were faster to judge whether the second number of a pair was larger or smaller than 5 if the previous number was numerically identical, pointing to a priming effect. Critically, this facilitation effect was found for both unimodal and cross-format number pairs.

Electrophysiologically, both the unimodal and the cross-format conditions elicited similar components. However, there were distinct effects of numerical identity in the unimodal and the cross-format conditions: Unimodal effects of numerical identity were found in the early time window between 100 and 200 ms after stimulus onset, as well as in the later time window between 250 and 400 ms. This suggests that pairs of visual-Arabic digits are linked at two different stages: First, both digits are rapidly and automatically integrated (as indexed by the N1 component), and second, their numerical content is processed semantically (as indexed by the N400 component). In contrast, cross-format effects were only found in the later time window between 250 and 400 ms. This dissociation implies that while both conditions involved semantic processing of the number pairs, only the unimodal condition involved a rapid and automatic integration of both constituents of a number pair. In other words, our results show that cross-format integration of numerical content occurs less rapidly than within-format integration. Conversely, a previous study did report effects of numerical identity with pairs of non-symbolic quantities and visual-Arabic digits already in the N1 time window (Liu et al., 2018). Therefore, it could be reasoned that the integration of different forms of symbolic number (digits, number words) takes longer than the integration of non-symbolic and symbolic quantities. However, there are certain methodological pitfalls to consider: In the study by Liu et al. (2018), all participants performed a behavioral estimation task with hundreds of trials immediately before the passive ERP task. In this estimation task, they were asked to estimate the quantity of the non-symbolic stimuli and type in their answers as Arabic digits, thus possibly entailing subvocal rehearsal of the respective number words. As identical non-symbolic stimuli were utilized during the passive ERP task, this may have likely provoked the co-activation of the respective Arabic digits, as well as the verbalized number words during the task.

Our finding of a more pronounced N400 for numerically non-identical than identical number pairs is in line with previous numerical cognition studies investigating the ERP correlates of semantic incongruencies (e.g., Niedeggen et al., 1999; Szücs et al., 2007: Szücs and Soltész, 2010; Pinhas et al., 2014). The finding of an N400 effect of numerical identity in the cross-format condition supports the hypothesis that number words and digits are only indirectly linked *via* their underlying numerical meaning, as proposed by semantic models of transcoding (e.g., Power and Dal Martello, 1990; McCloskey, 1992). Thus, the ERP results strongly suggest that the constituents of a number pair are processed semantically.

We set out to investigate the link between digits and number words. In this experiment, we did not find any evidence for

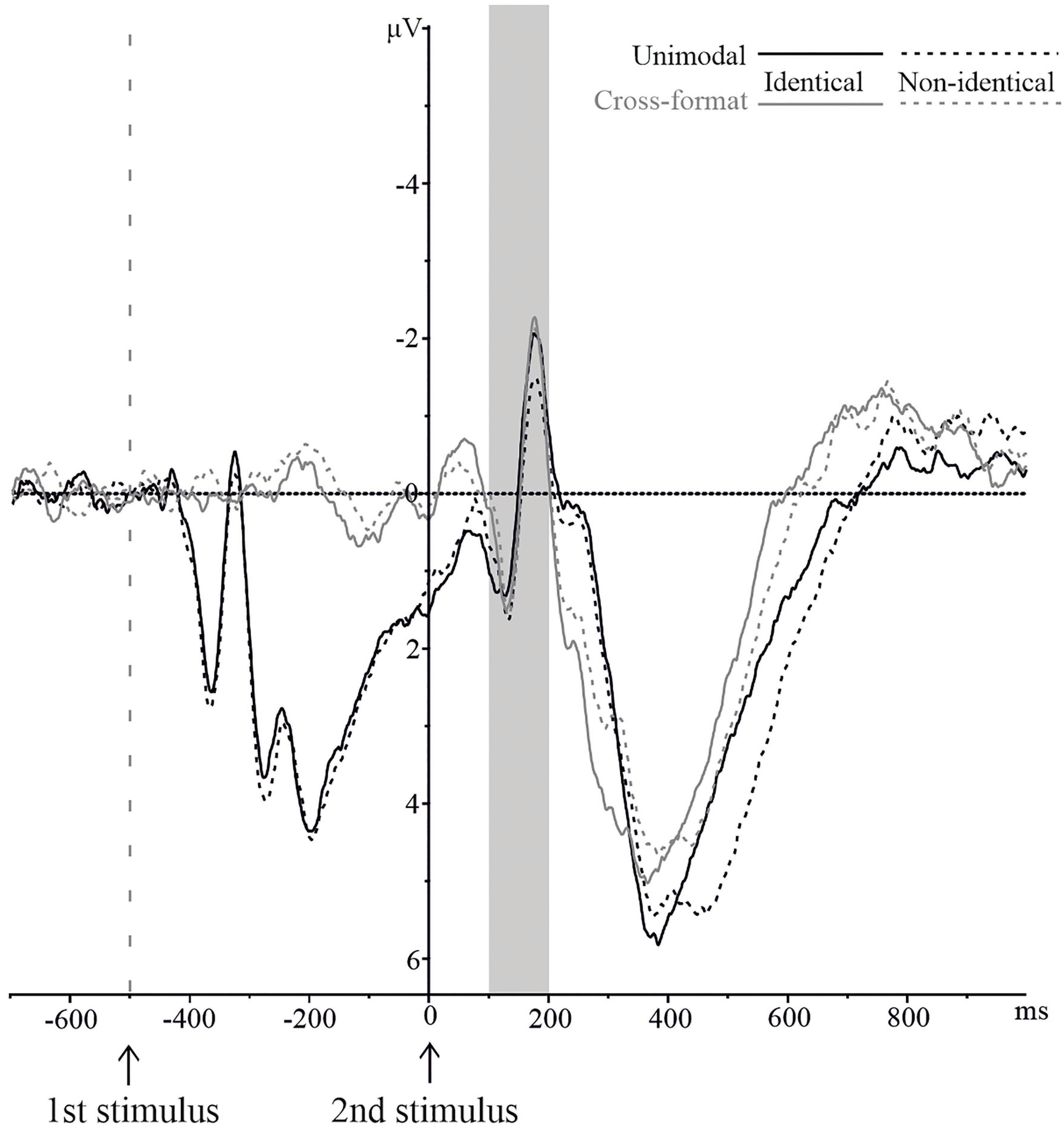


FIGURE 3 | N1 component on the pooled parietal electrode cluster for numerically identical and non-identical number pairs in unimodal visual and cross-format auditory–visual blocks. Solid lines represent ERPs for identical, and dashed lines for non-identical number pairs. ERPs are shown in black for unimodal items and in grey for cross-format items.

an early automatic cross-modal link. The presence of the later N400 effect of numerical identity for cross-format number pairs does suggest that representations of number words and digits were linked at a later time point. The N400 component is thought to reflect semantic processing; thus, our cross-format N400 effect suggests that number words and digits may be linked *via* number semantics. However, the task demands of Experiment 1 might have provoked an activation of semantic content, because participants were explicitly required to make numerical judgments. Arguably, participants may have actively and semantically processed the first number of a pair in order to facilitate the subsequent number judgment on the second number. This interpretation could also account for the behavioral priming effect we observed. It is also important to point out that our numerical judgment task involved response selection processes. This makes it difficult to distinguish whether the observed effects are due to numerical processing or response selection (Göbel et al., 2004).

In summary, we observed ERP effects of numerical identity for cross-format pairs of number words and visual-Arabic digits in Experiment 1. However, while unimodal pairs of visual-Arabic digits were associated with early N1 effects of

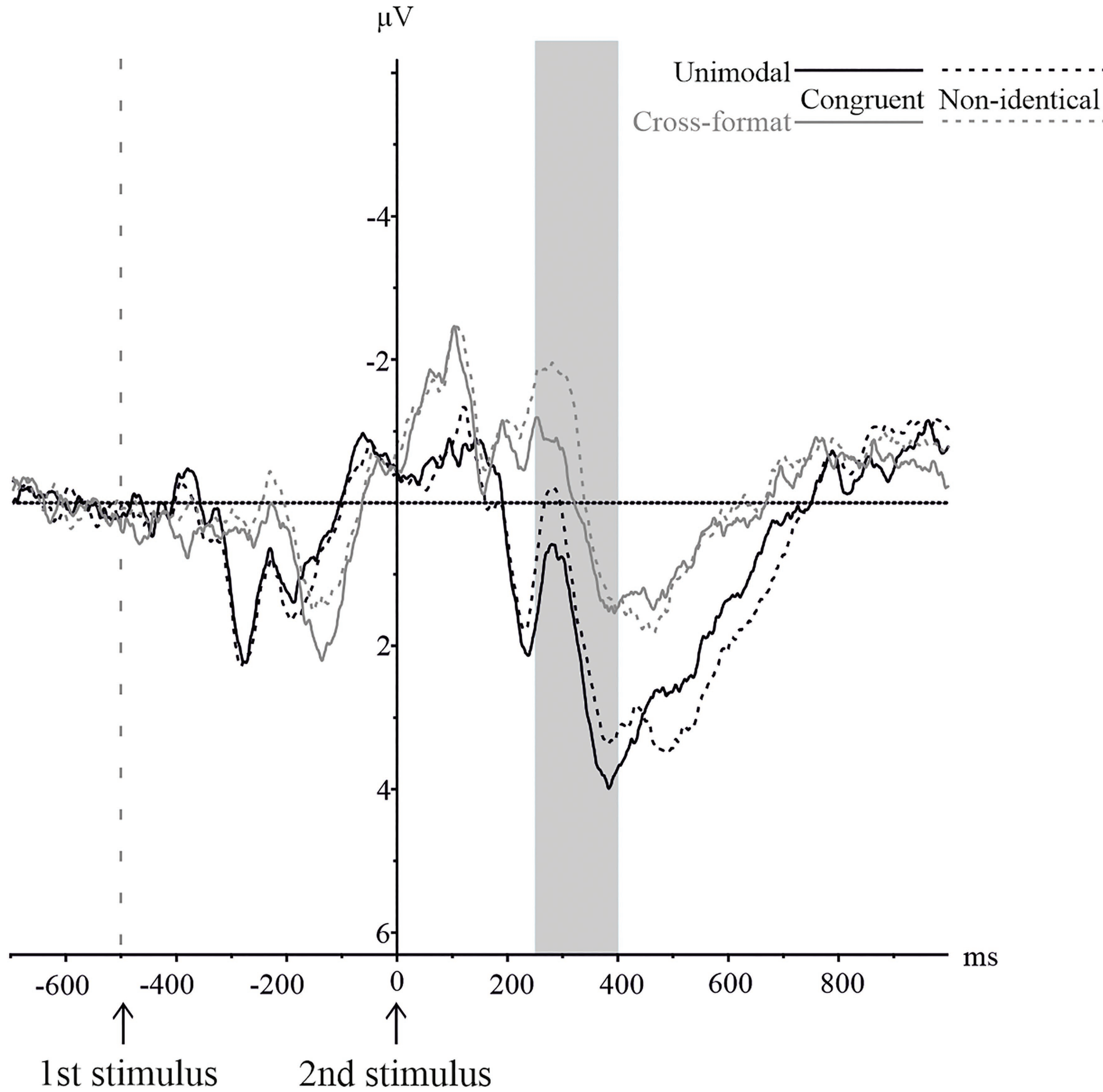


FIGURE 4 | N400 component on the pooled central electrode cluster for numerically identical and non-identical number pairs in unimodal visual and cross-format auditory–visual blocks. Solid lines represent ERPs for identical, and dashed lines for non-identical number pairs. ERPs are shown in black for unimodal items and in grey for cross-format items.

numerical identity (pointing to an automatic integration), cross-format numerical identity effects only emerged in the later N400 time window (pointing to semantic processing). In order to disentangle whether these cross-format N400 effects were due to the nature of the numerical judgment task, we performed another experiment in which participants were not explicitly required to access the underlying magnitude representation.

In Experiment 1, we had difficulties in finding the most suitable baseline correction. We had to settle on a period rather far away (−700 to −500 ms) from the onset of the target stimulus. As we can still expect amplitude changes due to the presentation of the first number after around 500 ms, we decided to increase the stimulus onset asynchrony (SOA) by 500 ms in Experiment 2.

EXPERIMENT 2

Participants

The sample comprised 19 healthy volunteers recruited at the University of Graz, Austria (age: $M = 25.2$ years, $SD = 3.1$; 8 males and 11 females). One additional participant had to be excluded from the data analysis because of noisy data. All participants were native speakers of German and had normal or corrected-to-normal vision, as well as normal hearing status. Participants received course credit or 10€ for participation. The study complied with the Declaration of Helsinki and approval was obtained from the ethics committee of University of Graz. Participants provided written informed consent prior to participation.

We conducted a power analysis to determine sample size "pwr" package (Champely, 2020) in R (R Core Team, 2020). We set power to 0.80 and the probability of alpha error to 0.05, corresponding to the convention by Cohen (1988). To obtain a conservative estimate, we decided to consider the smallest effect size of numerical identity found in Experiment 1, $\eta_p^2 = 0.242$. The power analysis revealed a minimum sample size of N = 10. Thus, sufficient power is guaranteed for our current sample of N = 19.

Stimuli and Procedure

Participants were presented with numbers and letters, some of which were moving, while others were stationary. Participants were instructed to indicate the movement direction for moving numbers, but not for moving letters *via* keypress. Unknown to the participants, stimuli were organized into 192 standard trials and 36 filler items per block. Standard trials consisted of a number pair, followed by a number moving horizontally across the screen. In half of the standard trials, the two numbers of a pair were identical (both numbers had the same numerosity, e.g., 1–1) or non-identical (the numbers were numerically different, e.g., 1–9). This sums up to a total of 96 identical and 96 non-identical trials per block. Stimuli consisted of digits and number words corresponding to the numerosities 1, 4, 6 and 9.

Note that the overt task (response to moving numbers) did not require participants to actively access the magnitude of the numbers. Importantly, the EEG analysis (see below) focused on the second number of the number pair preceding the moving number and not on the moving number itself, which makes our task a passive paradigm with respect to analysis of numerical congruency. Moreover, this design ensured that processing of the number pairs was not contaminated by eye movement artifacts caused by the moving numbers.

Since participants only had to respond after every third item, we wanted to make sure that they also had to actively attend to the first two items. Therefore, we included 12 filler items in which moving numbers appeared at the very beginning or after the presentation of just one number. To make sure that participants not only react to the perception of movement they were only asked to respond to moving numbers and not letters. Thus, we inserted 24 filler items with moving letters instead of numbers, which participants were instructed not to respond to. While the focus of Experiment 2 was to analyze numerical identity, additional factors were controlled to avoid predictive learning: Each first number was followed with the same probability by either the same or one specific different number (e.g., 1–1, 1–9). Also, number pairs did not predict the subsequent moving number: The moving number either had the same numerosity as the preceding number (25% of cases) or not (75% of cases). In half of the moving numbers, the movement direction and numerical size (a larger number, ≥6, moves to the right side of the screen or a smaller number, ≤4, moves to the left side of the screen) matched. In the other half, they did not (a larger number, ≥6, moves to the left side of the screen or a smaller number, ≤4, moves to the right side of the screen).

There were two experimental blocks: The unimodal block consisted only of visually presented digits, whereas in the cross-format block, the first number was always a spoken number word, while the second number and the moving number were visual-Arabic digits. At the beginning of the experiment, participants completed 12 practice trials with feedback. In the middle and at the end of each experimental block, participants had the opportunity to take a break. The order of the two blocks was counterbalanced.

Visual-Arabic stimuli were presented in white on black background with a height of 4 degrees of visual angle at the center of the display. Similar to Experiment 1, we controlled for low-level perceptual adaptation effects. For that reason, visual-Arabic numbers were displayed in one of four slightly different spatial locations at one degree of visual angle from the center of the display. The location followed a pre-defined pseudorandom order, in which no two stimuli appeared at the same location. In each trial, visual-Arabic numbers immediately following each other were displayed in different spatial locations. Number words were presented by one of four speakers (two male and two female voices). All number words had a duration of 500 ms.

As illustrated in **Figure 5**, each standard trial proceeded in the following order: Blank screen (500 ms), first number (500 ms), blank screen (500 ms), second number (500 ms), and blank screen (jitter: 400–600 ms). We analyzed ERPs in response to the second number. At the end of each trial, a digit moved horizontally to the left or the right side of the screen, until it stopped at a distance of four degrees of visual angle from the borders of the screen. The number moved at a constant speed of 1.67 pixels per frame. Participants were required to press the keyboard arrow corresponding to the direction of the movement either during the movement of the digit or after its arrival at the stationary position at the border of the screen. They were instructed to press the right arrow with their right index finger and the left arrow with their left index finger. If participants did not respond within 4 s of stimulus onset, the next trial was presented.

ERP Recording and Data Analysis

We employed the same ERP recording protocol as in Experiment 1, and the steps for preprocessing and data analysis were identical, except for baseline correction. The time window of −200 to 0 ms before onset of the second number of a pair served as the basis for baseline correction. Only segments with a correct response were considered. All participants had at least 74 valid segments in each of the four conditions; thus, all participants were included in the analyses. For the unimodal block, an average of 93.26 (*SD* = 3.87) identical and 92.32 (*SD* = 3.84) non-identical segments were retained. For the cross-format block, we kept 92.63 identical (*SD* = 5.35) and 92.26 non-identical (*SD* = 5.24) segments.

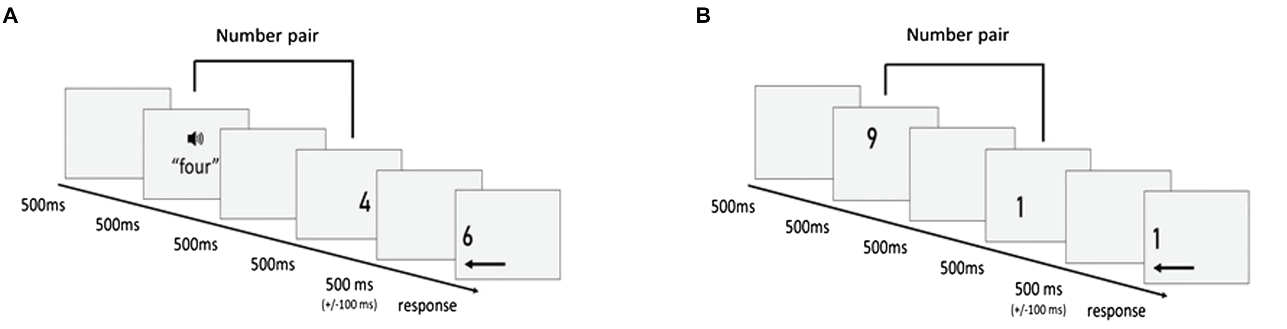


FIGURE 5 | Examples of different trials of the ERP paradigm: **(A)** unimodal visual block, non-identical number pairs, **(B)** cross-format auditory–visual block, identical number pairs.

Results

Behavioral Measures: Accuracy

To ensure that participants attended to the presented stimuli, they were required to respond to moving numbers, but not letters. On average, participants correctly reacted to 99.32% of the moving numbers (*SD* = 0.48%; range: 98.53–100.00%), while they incorrectly responded to only 9.58% of letters (*SD* = 6.63%, accuracy range: 65.96–97.87%). This high response accuracy suggests that participants were attentive toward the presented stimuli.

N1

As illustrated in **Figure 6**, the averaged waveforms of the parietal electrode cluster were more negative for non-identical than identical number pairs in the unimodal block, whereas this was not the case for the cross-format block. We performed an identity x modality ANOVA which showed that both main effects were not significant: modality, $F(1,18) = 4.02$, $p = 0.060$, $\eta_p^2 = 0.18$ and identity, $F(1,18) = 1.53$, $p = 0.231$, $\eta_p^2 = 0.08$. However, there was a significant interaction of modality x identity, $F(1,18) = 4.89$, $p = 0.040$, $\eta_p^2 = 0.21$.

To follow up on the significant interaction, we conducted two separate repeated measures ANOVAs with identity (identical vs. non-identical) as within-subject factor. These revealed a significant effect of identity for the unimodal block, $F(1,18) = 5.42$, $p = 0.032$, $\eta_p^2 = 0.23$, with more negative peak amplitudes for identical than non-identical number pairs. For the cross-format block, there was no significant difference, $F(1,18) = 0.03$, $p = 0.864$, $\eta_p^2 = 0.00$.

N400

The averaged waveforms of the central electrode clusters by numerical identity and experimental block are depicted in **Figure 7**. As illustrated in **Figure 7**, the averaged waveforms for non-identical number pairs were more negative than for identical number pairs, especially in the unimodal block. However, an identity x modality ANOVA showed no significant main effect of identity, $F(1,18) = 0.78$, $p = 0.388$, $\eta_p^2 = 0.04$. There was a significant main effect of modality, $F(1,18) = 9.95$, $p = 0.005$, $\eta_p^2 = 0.36$, with more negative peak amplitudes in the cross-format than in the unimodal block. The interaction identity x modality was not significant, $F(1,18) = 0.443$, $p = 0.514$, $\eta_p^2 = 0.024$.

Because the sample size was relatively small, we also conducted a Bayes factor (BF) analysis to determine the relative strength of the alternative hypothesis compared to the null hypothesis for the N400 ERP peak amplitude data (Dienes, 2014; Wagenmakers et al., 2018). We used the JASP software version 0.14.1.0 (JASP Team, 2020). As our repeated measures ANOVA contained several factors, we calculated inclusion Bayes factors ($BF_{Inclusion}$), which can be interpreted as evidence in the data for including a predictor (Wagenmakers et al., 2018). We found extreme evidence for the main effect of modality, $BF_{Inclusion} = 153.41$. For the main effect of identity, we found evidence for the null hypothesis, $BF_{Inclusion} = 0.30$. For the interaction, we found no evidence for the alternate hypothesis, $BF_{Inclusion} = 0.36$.

Discussion

The results of Experiment 2 support the notion that number pairs and digits are not automatically linked when no numerical judgments are involved. While presenting both a condition with unimodal pairs of visual-Arabic digits and cross-format pairs of number words and digits similar to Experiment 1, the present task was designed to be passive and to not require semantic number activation.

Unimodally, we found an early N1 effect of numerical identity and some traces of an N400 effect. Although the cross-format condition elicited similar components, these were not affected by numerical identity. The dissociation in the N1 component points again to an automatic integration of unimodal pairs of digits, but not of cross-format pairs of digits and number words. Unimodal integration of numerical stimuli therefore appears to happen automatically and involuntarily, even in a task not requiring any link between both constituents of a number pair. This supports the suggestion that processing of numerical identity is not limited to situations in which numerical information is explicitly processed (Liu et al., 2018). However, this automatic integration does not appear to extend beyond the visual modality, as we did not find any evidence for automatic integration of numerical

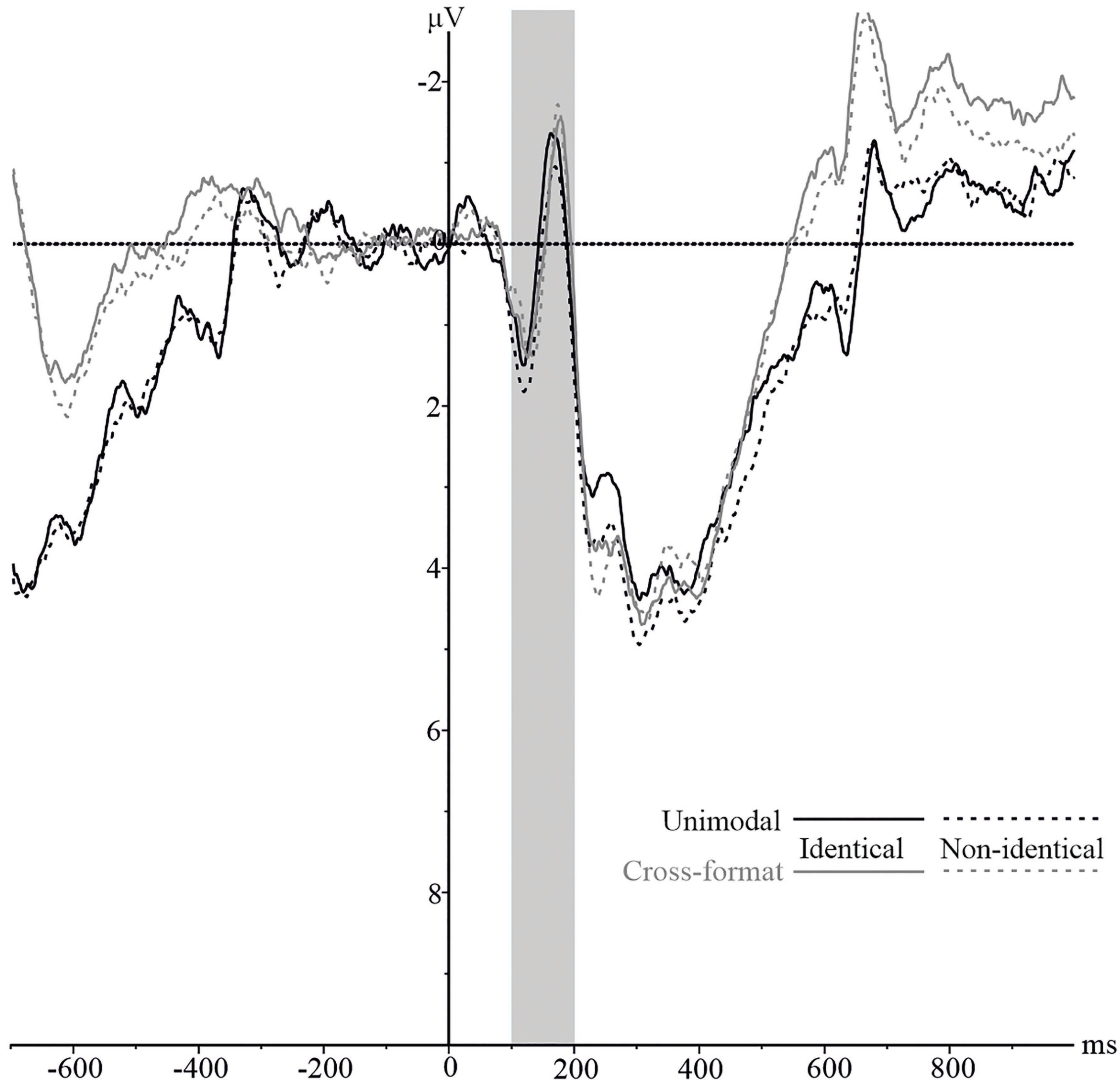


FIGURE 6 | N1 component on the pooled parietal electrode cluster for numerically identical and non-identical number pairs in unimodal visual and cross-format auditory–visual blocks. Solid lines represent ERPs for identical, and dashed lines for non-identical number pairs. ERPs are shown in black for unimodal items and in grey for cross-format items.

information from different sensory modalities. This corroborates findings from a previous study on visual–auditory integration of number words and non-symbolic quantities in preschoolers, which also did not report any early signs of integration (Pinhas et al., 2014).

The N400 effect which we found in Experiment 1 was basically eliminated by employing a task in which semantic activation of the underlying magnitude was not provoked. This is in contrast to the study by Pinhas et al. (2014), which reported higher N400 amplitudes for numerically non-identical than identical pairs of visually presented non-symbolic quantities and spoken number words. Arguably, the link between number words and their non-symbolic counterparts might be tighter than between number words and digits, at least in children. Further studies are necessary before drawing the conclusion that there is no automatic link between number words and digits in the absence of semantic processing.

GENERAL DISCUSSION

We investigated the possibly automatic link between number words and digits and examined whether unimodal numerical identity is associated with different ERP effects compared to cross-format numerical identity. We were interested in two ERP components: the early N1 component which is associated with automatic processing (Liu et al., 2018) and the later N400 component which is associated with the semantic processing of numbers (Niedeggen et al., 1999; Galfano et al., 2004; Paulsen and Neville, 2008; Szücs and Soltész, 2010).

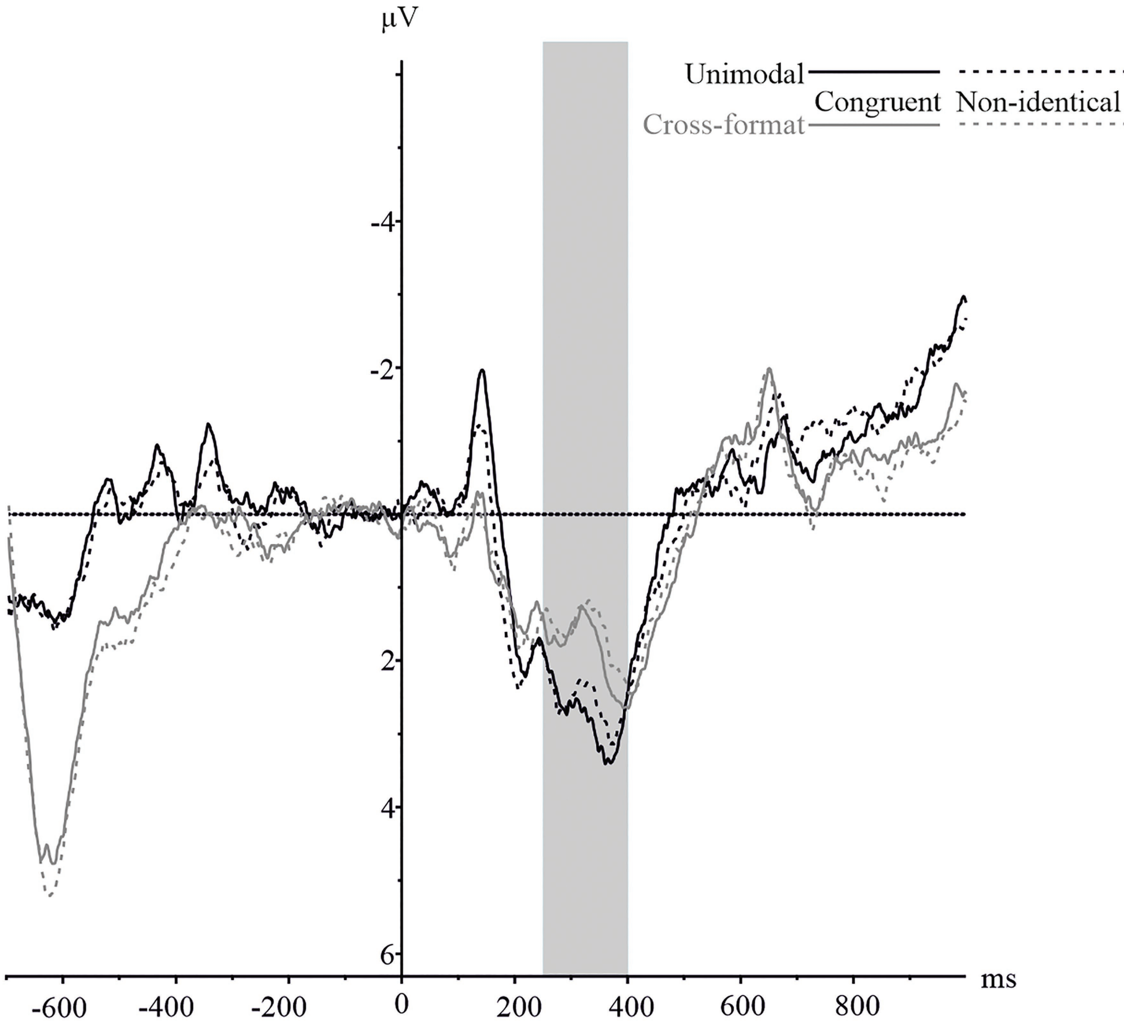


FIGURE 7 | N400 component on the pooled central electrode cluster for numerically identical and non-identical number pairs in unimodal visual and cross-format auditory–visual blocks. Solid lines represent ERPs for identical, and dashed lines for non-identical number pairs. ERPs are shown in black for unimodal items and in grey for cross-format items.

We found parietal N1 ERP effects of numerical identity in the early time window from 100 to 200 ms after stimulus onset. However, this was only found for unimodal, but not cross-format number pairs. On the one hand, this implies that visual-Arabic digits are rapidly and automatically linked, even if numerical processing is not actively required. On the other hand, automatic integration does not appear to extend to cross-format pairs of digits and number words. This suggests that cross-format integration of numerical information from different symbolic formats occurs less rapidly than within-format integration. This contrasts with previous research on the cross-format integration of visually presented non-symbolic and symbolic numerosities, which appear to be automatically linked (Liu et al., 2018). Therefore, symbolic numbers and their non-symbolic counterparts may have a tighter link than different symbolic representations.

Moreover, it is possible that our finding of an automatic integration of two visual-Arabic digits is partly due to perceptual visual similarity (c.f. Cohen, 2009) seeing that numerically identical number pairs showed a greater visual overlap than numerically non-identical ones. Evidence from an fMRI adaptation study with visual-Arabic and Chinese numerals suggests that brain responses to numerical stimuli are not only based on the numerical meaning but are also influenced by perceptual overlap (Holloway et al., 2013). Nonetheless, we did take measures to decrease this influence by varying the spatial locations and fonts of pairs of visual-Arabic digits. We found unimodal and cross-format N400 effects of numerical identity, but only when the active task required numerical decisions. As the N400 component is believed to reflect semantic processing (Kutas and Federmeier, 2011), it can be deduced that a semantic link between two numbers is not established directly, but instead individuals actively have to access the underlying meaning. While we did not find any evidence for an automatic link between cross-format number pairs as indexed by the N1 component, we did find an ERP effect of numerical identity for cross-format number pairs in the later N400 window.

Crucially, this was only true when an activation of semantic content was provoked by task demands (Experiment 1). When semantic activation was not provoked by the task (Experiment 2), we could not observe any N400 effects, neither unimodally nor in the cross-format condition. This suggests that number words and digits are indirectly linked *via* their underlying numerical magnitude. However, it is important to note that the SOA between the constituents of a number pair varied between Experiment 1 and Experiment 2: While the SOA was 500 ms in Experiment 1, it was 1,000 ms in Experiment 2. Therefore, it is important to consider the possibility that this longer SOA caused semantic priming to fade away and dissipate (e.g., Xiao and Yamauchi, 2017). First evidence suggests that this may indeed be the case in the domain of numerical processing: Lin and Göbel (2019) conducted a behavioral study in which participants were asked to indicate whether cross-format pairs of visual-Arabic digits and auditory number words with varying SOAs (−500 ms to +500 ms) were identical or not. Lin and Göbel (2019) observed cross-format numerical distance effects (indicating semantic processing) across all SOAs, but these distance effects decreased with increasing SOAs. Therefore, it is possible that the lack of semantic priming we observed in Experiment 2 (with an SOA of 1,000 ms) may have been related to using a long SOA. However, the precise effect of SOA length on semantic priming is still a matter of debate: Other studies suggest that semantic priming of lexical content is facilitated by longer SOAs (i.e., longer than 500 ms, e.g., Chen and Spence, 2018; Roelke et al., 2018). If this were also the case for cross-format number pairs, the lack of semantic priming we observed in Experiment 2 may not have been observed because, but perhaps rather in spite of a long SOA.

As discussed above, the current study found no evidence for a direct link between visual-Arabic numbers and number words. These results can easily be integrated with semantic models of transcoding, as they suggest accessing the other form through semantic activation. Asemantic models, however, are based on the assumption that transcoding takes place in the absence of semantic activation. Indeed, there is neuropsychological evidence supporting the view that semantic activation is not mandatory for transcoding: Dehaene and Cohen, 1995 described the case of a patient with Gerstmann's syndrome who was selectively impaired in tasks requiring access to the number semantics but showed intact transcoding skills. Nonetheless, the precise cognitive mechanisms linking number words and digits remain as yet unclear. A crucial step in support of asemantic models would be to demonstrate the existence of an automatic integration of number words and digits. The current study was, however, unable to do so in healthy adults.

An analogy for the absence of an automatic link between symbolic representations of number (i.e., number words and digits) can be found in the neighboring domain of reading. As a cautionary note, it is important to mention that there are distinctive differences between reading and number processing: While numbers are inherently meaningful as they reflect non-symbolic quantities, letters often have to be grouped to strings to form meaningful words. However, there are interesting parallels, as both letters and digits are culturally acquired symbols: While we communicate about quantities with number words and digits, our script code consists of strings of letters or characters that are used to reflect speech sounds. Evidence from cross-script priming suggests that there is indeed no automatic link between different scripts within the same language (Okano et al., 2013). Specifically, cross-script priming can be investigated in languages containing pairs of symbols that map onto the same phonological representation. For instance, Japanese has two syllabaries, Hiragana and Katakana, which both have characters directly corresponding to the same Japanese syllables (e.g., Hiragana さ and Katakana サ both represent/sa/). Similar to our behavioral findings on the cross-format priming of number words and digits, substantial behavioral priming effects for primes that are displayed in different scripts from their targets have been reported both in lexical decision (Pylkkänen and Okano, 2010) and semantic categorization tasks (Okano et al., 2013). However, ERP-based findings suggested that cross-script prime-target pairs are not automatically linked, but rather *via* their underlying semantics as indexed by the N400 component (Okano et al., 2013).

Interestingly, similar to cross-script priming, there is a body of evidence supporting the notion of an automatic link between non-symbolic quantities and their symbolic counterparts (Galfano et al., 2004; Paulsen and Neville, 2008; Pinhas et al., 2014; Liu et al., 2018). It would be fruitful to disentangle possibly different mechanisms supporting the integration of non-symbolic visual quantities and number words, compared to symbolic Arabic digits and number words. A future challenge for numerical cognition research therefore is to investigate whether the automatic integration of non-symbolic quantities and their symbolic counterparts contributes to higher-order skills such as mental calculation.

It is important to acknowledge some limitations of the current study. One might argue that the observed absence of an automatic and asemantic integration of auditory number words and visual-Arabic digits may be partially due to our experimental design. First, early ERP components such as the N1 effect are known to be modality-dependent (Donohue et al., 2011), and second, the length of the SOAs between cross-format number pairs may have impacted the automatic association (Lin and Göbel, 2019).

Concerning modality-dependence, it is possible that the observed N1 effects of numerical identity constitute a unique feature of unimodal processing of visually presented numbers. This may also explain why Liu et al. (2018) observed early ERP effects for the integration of visually presented quantities and digits, while the present study could not find any early signs of cross-format integration between auditory number words and visual-Arabic digits. However, we compared ERPs within and not across modalities. Indeed, for both unimodal and cross-format conditions, we compared the N1 effect evoked by the presentation of the second number of a pair, which was always presented visually (i.e., unimodal: two visual-Arabic digits; cross-format: one auditory number word followed by one visual-Arabic digit). Since no cross-format comparisons were made, the observed similarities and differences should not stem from modality-specific effects.

A central parameter in early cross-modal integration is the temporal proximity of the stimuli (e.g., Donohue et al., 2011). The shorter the SOA, the stronger the observed priming effect (Lin and Göbel, 2019). Short SOAs, however, do not enable the separate analysis of ERPs, as high amplitude components of the prime stimulus may still take place during the time window of the target stimulus. To keep the priming effect as large as possible while keeping the ERPs of the first and second stimuli as dissociated as possible, we employed a 500-ms delay in Experiment 1 and a 1,000-ms delay in Experiment 2. We managed to observe unimodal identity effects with such a design and assume that if there were automatic markers of cross-modal integration, we should have been able to detect those with our design. However, while we used different SOAs in Experiments 1 and 2, our results cannot directly inform the question about the effect of manipulating SOAs because of additional design differences between the two experiments. Such effects should be examined in future experiments only manipulating SOAs.

CONCLUSION

Our findings contribute to the debate on the nature of the integration of different symbolic number forms (visual-Arabic digits and auditory number words). In both experiments, unimodal pairs of visual-Arabic digits were consistently found to be automatically integrated across both experiments, but we did not find any evidence for an early and automatic cross-format integration. In our experiments, evidence of the cross-format association between visual-Arabic digits and verbal number words emerged late and involved semantic activation. The present study thus does not support the notion of an automatic and asemantic cross-format integration of number words and visual-Arabic digits in adults.

DATA AVAILABILITY STATEMENT

The data collected for this study is publicly available on the Open Science Framework and can be accessed at: https://osf.io/p7ksn/.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

KL and SG developed the project concept. SF, FK, CB, FC, AS, and SG developed ideas for the design of the experiments. SF programmed the experiments, and analyzed and interpreted the data together with FK. SF, CP-S, CB, and AS acquired the data, and IP provided the required resources. SF drafted the manuscript under the supervision of FK and KL, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

FUNDING

This work was supported by the FWF, Austria (grant no. I 2778-G16) and the ESRC, UK (grant no. ES/N014677/1).

ACKNOWLEDGMENTS

We would like to thank Theresia Volk and Viktoria Wirth for helping with data collection.

REFERENCES

Avancini, C., Galfano, G., and Szücs, D. (2014). Dissociation between arithmetic relatedness and distance effects is modulated by task properties: An ERP study comparing explicit vs. implicit arithmetic processing. *Biol. Psychol.* 103, 305–316. doi: 10.1016/j.biopsycho.2014.10.003

Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: A developmental, asemantic, and procedural model for transcoding from verbal to Arabic numerals. *Psychol. Rev.* 111, 368–394. doi: 10.1037/0033-295X.111.2.368

Bassok, M., Fisher, K., and Osterhout, L. (2009). "Conceptual Integration in Arithmetic is the Same for Digits and for Words: It's the Meaning, Stupid!" in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2142–2147.

Champely, S. (2020). Basic Functions for Power Analysis (R package Version 1.3–0; Computer software). Available at: https://cran.r-project.org/package=pwr (Accessed August 28, 2021).

Chen, Y. C., and Spence, C. (2018). Dissociating the time courses of the cross-modal semantic priming effects elicited by naturalistic sounds and spoken words. *Psychon. Bull. Rev.* 25, 1138–1146. doi: 10.3758/s13423-017-1324-6

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences Lawrence Erlbaum Associates.

Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychon. Bull. Rev.* 16, 332–336. doi: 10.3758/PBR.16.2.332

Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120. doi: 10.3758/PBR.16.2.332. Integers

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Front. Psychol.* 5:781. doi: 10.3389/fpsyg.2014.00781

Donohue, S. E., Roberts, K. C., Grent-'t-Jong, T., and Woldorff, M. G. (2011). The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. *J. Neurosci.* 31, 7982–7990. doi: 10.1523/JNEUROSCI.5298-10.2011

Dotan, D., and Friedmann, N. (2018). A cognitive model for multidigit number reading: inferences from individuals with selective impairments. *Cortex* 101, 249–281. doi: 10.1016/j.cortex.2017.10.025

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

Galfano, G., Mazza, V., Angrilli, A., and Umiltà, C. (2004). Electrophysiological correlates of stimulus-driven multiplication facts retrieval. *Neuropsychologia* 42, 1370–1382. doi: 10.1016/j.neuropsychologia.2004.02.010

Galfano, G., Penolazzi, B., Vervaeck, I., Angrilli, A., and Umiltà, C. (2009). Event-related brain potentials uncover activation dynamics in the lexicon of multiplication facts. *Cortex* 45, 1167–1177. doi: 10.1016/j.cortex.2008.09.003

Göbel, S. M., Johansen-Berg, H., Behrens, T., and Rushworth, M. F. S. (2004). Response-selection-related parietal activation during number comparison. *J. Cogn. Neurosci.* 16, 1536–1551. doi: 10.1162/0898929042568442

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

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9

Guthormsen, A. M., Bassok, M., and Osterhout, L. (2017). Knowledge: evidence from event-related potentials. *Cogn. Sci.* 40, 723–757. doi: 10.1111/cogs.12238. Conceptual

Hagemann, D. (2004). Individual differences in anterior EEG asymmetry: methodological problems and solutions. *Biol. Psychol.* 67, 157–182. doi: 10.1016/j.biopsycho.2004.03.006

Holloway, I. D., Battista, C., Vogel, S. E., and Ansari, D. (2013). Semantic and perceptual processing of number symbols: evidence from a cross-linguistic fMRI adaptation study. *J. Cogn. Neurosci.* 25, 388–400. doi: 10.1162/jocn_a_00323

Hyde, D. C., and Spelke, E. S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number rapresentations. *J. Cogn. Neurosci.* 21, 1039–1053. doi: 10.1162/jocn.2009.21090

JASP Team. (2020). JASP 0.14.1.0 (Computer software). https://jasp-stats.org/

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Libertus, M. E., Woldorff, M. G., and Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behav. Brain Funct.* 3, 1–15. doi: 10.1186/1744-9081-3-1

Lin, C. Y., and Göbel, S. M. (2019). Arabic digits and spoken number words: timing modulates the cross-modal numerical distance effect. *Q. J. Exp. Psychol.* 72, 2632–2646. doi: 10.1177/1747021819854444

Liu, R., Schunn, C. D., Fiez, J. A., and Libertus, M. E. (2018). The integration between nonsymbolic and symbolic numbers: evidence from an EEG study. *Brain Behav.* 8:e00938. doi: 10.1002/brb3.938

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., 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

Malone, S. A., Burgoyne, K., and Hulme, C. (2020). Number knowledge and the approximate number system are two critical foundations for early arithmetic development. *J. Educ. Psychol.* 112, 1167–1182. doi: 10.1037/edu0000426

Marinova, M., Sasanguie, D., and Reynvoet, B. (2018). Symbolic estrangement or symbolic integration of numerals with quantities: methodological pitfalls and a possible solution. *PLoS One* 13:e0200808. doi: 10.1371/journal.pone.0200808

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. *Cognition* 44, 107–157. doi: 10.1016/0010-0277(92)90052-J

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

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2015). From "five" to 5 for 5 minutes: Arabic number transcoding as a short, specific, and sensitive screening tool for mathematics learning difficulties. *Arch. Clin. Neuropsychol.* 30, 88–98. doi: 10.1093/arclin/acu071

Niedeggen, M., Rösler, F., and Jost, K. (1999). Processing of incongruous mental calculation problems: evidence for an arithmetic N400 effect. *Psychophysiology* 36, 307–324. doi: 10.1017/S0048577299980149

Okano, K., Grainger, J., and Holcomb, P. J. (2013). An ERP investigation of visual word recognition in syllabary scripts. *Cogn. Affect. Behav. Neurosci.* 13, 390–404. doi: 10.3758/s13415-013-0149-7

Paulsen, D. J., and Neville, H. J. (2008). The processing of non-symbolic numerical magnitudes as indexed by ERPs. *Neuropsychologia* 46, 2532–2544. doi: 10.1016/j.neuropsychologia.2008.04.003

Peirce, J. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform*. 2:10. doi: 10.3389/neuro.11.010.2008

Pinhas, M., Donohue, S. E., Woldorff, M. G., and Brannon, E. M. (2014). Electrophysiological evidence for the involvement of the approximate number system in preschoolers' processing of spoken number words. *J. Cogn. Neuro.* 26, 1891–1904. doi: 10.1162/jocn_a_00631

Power, R. J. D., and Dal Martello, M. F. (1990). Language and cognitive processes The dictation of Italian numerals The dictation of Italian numerals. *Lang. Cogn. Neurosci.* 5, 237–254. doi: 10.1080/01690969008402106

Power, R. J. D., and Dal Martello, M. F. (1997). From 834 to eighty thirty four: The Reading of Arabic numerals by seven-year-old children. *Math. Cogn.* 3, 63–85. doi: 10.1080/135467997387489

Purpura, D. J., Baroody, A. J., and Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: mediation by numeral knowledge. *J. Educ. Psychol.* 105, 453–464. doi: 10.1037/a0031753

Pylkkänen, L., and Okano, K. (2010). The nature of abstract orthographic codes: evidence from masked priming and magnetoencephalography. *PLoS One* 5:e10793. doi: 10.1371/journal.pone.0010793

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., and Hofmann, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychon. Bull. Rev.* 25, 1488–1493. doi: 10.3758/s13423-018-1453-6

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

Sasanguie, D., and Reynvoet, B. (2014). Adults' arithmetic builds on fast and automatic processing of Arabic digits: evidence from an audiovisual matching paradigm. *PLoS One* 9, 1–6. doi: 10.1371/journal.pone.0087739

Szücs, D., and Csépe, V. (2005). The effect of numerical distance and stimulus probability on ERP components elicited by numerical incongruencies in mental addition. *Cogn. Brain Res.* 22, 289–300. doi: 10.1016/j.cogbrainres.2004.04.010

Szücs, D., and Soltész, F. (2010). Event-related brain potentials to violations of arithmetic syntax represented by place value structure. *Biol. Psychol.* 84, 354–367. doi: 10.1016/j.biopsycho.2010.04.002

Szücs, D., Soltész, F., Czigler, I., and Csépe, V. (2007). Electroencephalography effects to semantic and non-semantic mismatch in properties of visually presented single-characters: The N2b and the N400. *Neurosci. Lett.* 412, 18–23. doi: 10.1016/j.neulet.2006.08.090

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

Vogel, S. E., Goffin, C., Bohnenberger, J., Koschutnig, K., Reishofer, G., Grabner, R. H., et al. (2017). The left intraparietal sulcus adapts to symbolic number in both the visual and auditory modalities: evidence from fMRI. *NeuroImage* 153, 16–27. doi: 10.1016/j.neuroimage.2017.03.048

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2018). Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychol. Bull. Rev.* 25, 35–57. doi: 10.3758/s13423-017-1343-3

Xiao, K., and Yamauchi, T. (2017). The role of attention in subliminal semantic processing: A mouse tracking study. *PLoS One* 12, 1–17. doi: 10.1371/journal.pone.0178740

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Education

ORIGINAL RESEARCH
published: 18 January 2022
doi: 10.3389/feduc.2021.642153

Check for updates

Place Value Understanding Explains Individual Differences in Writing Numbers in Second and Third Graders But Goes Beyond

Moritz Herzog[1]* *and Annemarie Fritz*[2]

[1]Institute of Educational Research, School of Education, University of Wuppertal, Wuppertal, Germany, [2]Akademie Wort + Zahl. Internationales Institut für Interkulturelle Bildung, Cologne, Germany

Recent studies have shown that children's proficiency in writing numbers as part of the so-called transcoding correlates with math skills. Typically, children learn to write numbers up to 10,000 between Grade 1 and 4. Transcoding errors can be categorized in lexical and syntactical errors. Number writing is thus considered a central aspect of place value understanding. Children's place value understanding can be structured by a hierarchical model that distinguishes five levels. The current study investigates to what extent a profound understanding of the place value system can explain individual differences in number writing. N = 266 s and third graders (126 girls) participated in the study. The children wrote down 28 verbal given numbers up to 10,000 and completed a place value test based on a hierarchical model to assess number writing skills and place value understanding. Second graders made more number writing errors than third graders and transcoding errors were mostly syntactical errors. In both grades, transcoding performance and place value understanding correlated substantially. In particular complex numbers were more often solved correctly by children with a more elaborated place value understanding. The effect of place value understanding on error rate was smaller regarding lexical errors than syntactical errors. This effect was also comparably small regarding inversion-related errors. The results underpin that writing numbers is an integral part of early place value understanding. Writing numbers can be assumed to be mostly based on the identification of the place values. However, variance in transcoding skills cannot totally be explained by place value understanding, because children with an elaborated place value understanding differed in transcoding performance, too. The differences between the grades indicate that children's development of writing numbers is also driven by instruction in school. Thus, writing numbers and place value understanding overlap but exceed each other. We discuss how an understanding of the place value relations can be integrated in existing frameworks of place value processing. Since writing numbers is a basic skill in place value understanding, it might serve as an efficient screening method for children, who struggle severely with understanding the decimal place value system.

Keywords: place value understanding, transcoding, numerical development, numerical cognition, number writing

Edited by:
Ricardo Moura,
University of Brasilia, Brazil

Reviewed by:
Ren Liu,
University of California, Merced, United States
Ricardo Garcia,
Federal University of Paraíba, Brazil

***Correspondence:**
Moritz Herzog
mherzog@uni-wuppertal.de

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

Received: *15 December 2020*
Accepted: *13 December 2021*
Published: *18 January 2022*

Citation:
Herzog M and Fritz A (2022) Place Value Understanding Explains Individual Differences in Writing Numbers in Second and Third Graders But Goes Beyond. Front. Educ. 6:642153. doi: 10.3389/feduc.2021.642153

INTRODUCTION

When someone tells us his or her phone number, when we write down a friend's new address, or when we make a note to take the right bus line whose number a helpful stranger told us: Many every-day contexts require to write a number from verbal information. In research, the skill to write down numbers given verbally is often referred to as transcoding (Barrouillet et al., 2004; Gilmore et al., 2018). Transcoding, as the term indicates interrelates several codes of numbers, i.e., representations, in which numbers can appear.

Dehaene (1992) proposed three codes for number representations in the frequently cited *Triple Code Model*. The model consists of three codes of numbers that are interrelated: 1) a verbal system, which mostly refers to number words, but also verbally stored arithmetic facts in the log-term memory (e.g., multiplication table); 2) the visual system including Arabic numerals; and 3) a quantity system covering nonverbal number representations, such as sets of dots or positions on a number line (Dehaene, 1992; Dehaene and Cohen, 1995). Thus, a number can be represented in these three codes as well as transcoded between them.

The importance of transcoding abilities for mathematical learning is emphasized by its relation to mathematical performance. Empirical studies have shown that transcoding correlates with arithmetic performance (Geary et al., 1999; Moeller et al., 2011a; Göbel et al., 2014b). Moreover, number line estimation accuracy also correlates with transcoding (Dietrich et al., 2016). In line with these results, persistent transcoding difficulties are typical for children with mathematical learning difficulties (Geary et al., 1999; Moura et al., 2013; Moura et al., 2015). Especially numbers of higher complexity challenge children with mathematical learning difficulties even at the end of primary school, when their typical developing peers have mastered transcoding (Mark and Dowker, 2015; Moura et al., 2015; Houdement and Tempier, 2019).

Transcoding abilities usually develop during primary school. While many children at the beginning of primary school show difficulties with reading or writing numbers, most of them master transcoding by end of Grade 4 (Byrge et al., 2014; Moura et al., 2015). The difficulty of number reading and writing depends–among others–on the complexity of the numbers. Single-digit numbers are only slightly affected by transcoding errors, which appear mostly in multi-digit numbers (Zuber et al., 2009; Moura et al., 2015). Moura et al. (2015) have shown that children's transcoding proficiency regarding numbers of different complexity develops parallelly during the course of primary school: For example, first graders showed similar difficulties with numbers of low difficulty (e.g., 190) as second graders showed with numbers of moderate difficulty (e.g., 109).

The difficulty of number reading and writing is also affected by the number word system of a language. Number words vary in their transparency in different languages. For example, 35 (thirty-five) is read "sān shí wǔ" ("three ten five") in Chinese, which is more transparent than English as the tens are given in a decomposition (three ten) and not a new word (thirty). In German, it is "fünfunddreissig" ("five and thirty"), which is even less transparent than English due to the inversion of tens and units (Zuber et al., 2009; Göbel et al., 2014a). Comparative studies between different languages have shown that transcoding is easier in terms of accuracy and reaction time in more transparent languages (Miller et al., 1995; Pixner et al., 2011; Dowker and Nuerk, 2016).

In particular the tens-units-inversion that is found for example in German and Dutch is a major challenge in transcoding (Pixner et al., 2011; Klein et al., 2013). In a comparison study regarding the influence of different characteristics of numbers (e.g., number size, pronunciation of number words, or difference between digits) on transcoding, van der Ven et al., (2017) showed that inversion was a significant predictor for transcoding difficulty. Imbo et al. (2014) compared French (non-inverted) and Dutch (inverted) speaking second-graders. Although overall transcoding performance did not differ significantly between the language groups, the Dutch speaking children had a nearly six times higher inversion error rate than their French peers.

Empirical evidence suggests that working memory is involved in transcoding difficulties in inverted languages (Camos, 2008; Zuber et al., 2009; Pixner et al., 2011). Zuber et al. (2009) have shown that transcoding correlates with visual-spatial working memory and central executive, but not with phonological working memory in first graders. However, the central executive was only involved in inversion-related transcoding errors (e.g., 53 for 35), which indicates that coordinating number word parts for tens and units is a main difficulty during transcoding in languages with tens-units inversion. In line with that, Poncin et al., (2020) recently compared transcoding performance in inverted ("five and thirty") and non-inverted ("thirty-five") number words at the end of primary school. The French-speaking children could solve the transcoding task given in the non-inverted (usual in French) condition in significant shorter time than the inverted (unusual in French) condition. However, German speaking children were as fast in transcoding when presented inverted number words (usual in German) as when presented non-inverted number words (unusual in German). Obviously, the tens-units-inversion leads to increased reaction times in transcoding even in children who are used to it. This result highlights the cognitive cost of inverted number words. Lopes-Silva et al., (2014) investigated the role of verbal skills beyond working memory regarding transcoding processes in non-inverted number words. In their study, phonemic awareness outran working memory capacities regarding the prediction of transcoding performance.

Transcoding Processes

Transcoding numbers from verbal to Arabic code requires an understanding of the decimal rules of number word structures (Deloche and Seron, 1982; Pixner et al., 2011): The parts of the number words (e.g., nine hundred fifty-one) have to be mapped to corresponding numerals (900, 50, 1), which need to be composed according to certain, language-specific rules (951). Barrouillet et al. (2004) proposed an often adopted *asemantic*, *developmental*, *and procedural model for transcoding* (ADAPT model) to specify the processes involved in transcoding. The

ADAPT model focuses on procedural rules on the one hand and on the other hand on the construction of a "lexicon" for multi-digit numbers. According to the ADAPT model, transcoding does not involve eliciting details such as the meaning of the digits, given that a corresponding entry in the numerical lexicon is available. If there is no entry, procedural rules have to be employed. In this process, the number word forms a digit string that contains information about the digit value and the positional value.

Within the ADAPT model, four types of procedural transcoding rules are differentiated. Some numbers (e.g., 11) or digits (e.g., 9 in 951) are derived from long-term memory (P1 rules). When the digit strings have to be assembled, their length has to be derived from the number words as indicated by keywords such as "hundred" or "thousand" (P2 and P3 rules). P1 and P2/P3 rules are combined, when the digit value is derived from long-term memory and the position from the keyword, as in "nine hundred" ("nine" = P1, "hundred" = P2). The structure of many number words in triplets (e.g., three hundred thirty-nine thousand two hundred eleven) facilitates transcoding up to one million with just two of these rules (P2 for three-digit and P3 for four-digit numbers) (Barrouillet et al., 2004; Van de Walle et al., 2016). Finally, the written number has to be checked for completeness. If there are gaps in the digit string, they have to be filled up with zeros (P4 rules).

Two main error types in transcoding are distinguished: lexical and syntactical errors. While errors in mapping the corresponding number to a digit (e.g., 941 for 951) are considered lexical errors, wrong compositions of the numerals (e.g., 90,051 for 951) are called syntactical errors (Barrouillet et al., 2004; Deloche and Seron, 1982). After comparing lexical and syntactical transcoding errors in children with and without mathematical learning difficulties in early and middle primary school, Moura et al. (2013) reported three main effects: First, syntactical error rates were higher than lexical error rates in all children. Second, children at the beginning of primary school made more transcoding errors than children in middle primary school. And third, typical developing children showed lower transcoding error rates than children with mathematical learning difficulties. In particular, lexical error rates were very low in typical developing children both in early and middle primary school. However, first-graders with mathematical learning difficulties showed considerably higher lexical error rates. In contrast to lexical errors, syntactical errors were generally found in both groups and both stages of primary school, accounting for the three main effects. Syntactical errors mostly affected three- and four-digit numbers, while single- and two-digit numbers challenged only very few children (Moura et al., 2013).

As proposed in the ADAPT model, employing transcoding rules strongly draws on procedural knowledge such as identifying place values. Procedural knowledge refers to *how* rules and procedures (e.g., the rules of the ADAPT model) are carried out. In contrast to procedural knowledge, conceptual knowledge covers the understanding of *why* these rules and procedures apply and which structures underly them (Hiebert and Lefevre, 1986). Procedural and conceptual knowledge can be applied regarding place value understanding, too. While procedural place value understanding refers to knowledge of the place values and how they can be composed to multi-digit numbers, conceptual place value understanding can be identified with the iterative relation of the bundling units (i.e., hundreds, tens, units, etc.): ten units can be unitized to one ten and so on (Houdement and Tempier, 2019; Van de Walle et al., 2016). Rittle-Johnson and Schneider (2015) emphasize that procedural and conceptual mathematical knowledge are intertwined. The interrelation between procedural and conceptual knowledge implies that procedural skills may have a conceptual basis on which they are acquired and employed. In the case of transcoding, there is little known about its conceptual foundations.

Place Value Understanding

Against the background of the ADAPT model, transcoding implies specific knowledge of place value understanding. For the reliable employment of the procedural rules of the ADAPT model, children need to recognize the decimal unit a digit represents. According to Nuerk et al. (2014), place value information is processed in three ways: First, place value identification refers to the correct finding and naming of digit positions in a multi-digit number. Therefore, this aspect can be identified primarily with transcoding services. Second, place value activation refers to the employment–consciously or unconsciously–of the numerical information of a decimal unit, for example in number comparison tasks. Third, place value computation describes the integration of place value information in arithmetic tasks. This taxonomy is particularly used in (neuro-) psychological studies on number processing (Nuerk et al., 2014; Bahnmueller et al., 2018).

Naturally, place value understanding has been addressed by researchers from (mathematical) education, too. Based on the notion that transcoding is mostly based on a conceptual understanding of the place value system, procedural and asemantic models such as the ADPAT models have been criticized (e.g., Geary, 2004; Desoete and Grégoire, 2006). However, a profound understanding of the decimal place value system covers both procedural and conceptual aspects such as writing and reading numbers and insight in the iterative relation of the bundling units (Fuson et al., 1997a; Van de Walle et al., 2016; Herzog et al., 2019; Houdement and Tempier, 2019).

To structure the development of place value understanding, Herzog et al. (2019) proposed a developmental model of place value understanding that distinguishes five levels. The levels build up on each other hierarchically. That means that the level hierarchy implies a relation of dependence and inclusion between the levels: First, children need the knowledge of lower levels to develop successive levels (dependence). Second, children who have developed a certain level are supposed to have developed the prior levels, too (inclusion) (Battista, 2011). The levels of the model are not distinct classes of place value understanding that suddenly change. Rather, higher levels are elaborations and advancements of lower levels (Clements and Sarama, 2004). By interacting with tasks and materials based on the place value system (e.g., multi-digit arithmetic, base-ten blocks, standard algorithms), children develop a more elaborated understanding of how numbers are composed of place values and how decimal bundling units are related. As children's place value understanding does not change suddenly, but gradually over time, the levels are interrelated in form of overlapping waves (Siegler and Alibali, 2005; Clements and Sarama, 2014). Each level can be described by typical ideas of the place value system, strategies used to solve tasks, and errors made by the children at the respective level. The gradual development of children along the levels implies that a child can be located at a certain level of which the child can solve most items; however, this child might also be able to solve single items of higher levels (e.g., by applying a specific strategy) and, vice versa, make single errors on items of lower levels (e.g., due to careless mistakes).

The model is theoretically based on a broad literature review and content analysis. The most relevant literature for the construction of the model are earlier models on place value understanding (Cobb and Wheatley, 1988; Ross, 1989; Fuson, et al., 1997b). The influence of the earlier models gets visible in the description of the levels below. However, the earlier models cover only two-digit numbers, in contrast to the model by Herzog et al. (2019). The significance of the relation between bigger bundling units such as hundreds and thousands is highlighted in the literature (Scherer and Moser Opitz, 2010; Houdement and Tempier, 2019). While the earlier models are based on classroom observations, single cases and empirical studies, there is little a posteriori evidence for their validity (Chan et al., 2014). To the best of our knowledge there are no longitudinal studies supporting the validity of the earlier models for describing learning trajectories. Longitudinal studies are especially necessary for developmental models that describe children's typical learning trajectories (Reiss and Obersteiner, 2019). The content analysis of the place value system during the construction of the model stressed the relevance of both the identification of the place values (Nuerk et al., 2015) and the relation of the bundling units (Houdement and Tempier, 2019).

The model construction followed a four-step circle as suggested by Battista (2011). The first step was the literature review and content analysis described above. The second step was the construction of a provisory model and designing items corresponding to the levels of the provisory level. In a third step, the model was tested empirically in several piloting studies. Step four was the evaluation of the provisory model and the empirical results. In the few cases where the provisory model and the empirical findings of the piloting studies did not match, the model was slightly revised and tested again. The adaptation process of the model involved only few items that were carefully adapted to the revised models.

The final model was validated in a cross-sectional and a longitudinal study in Germany employing a one-dimensional Rasch-Analysis with a new sample (Herzog and Fritz, 2019). In the cross-sectional study, the item difficulties of the Rasch-Analysis followed the predictions of the level hierarchy in two ways: First, items operationalizing the same level showed similar difficulties; second, items of lower levels were systematically easier than items of higher levels. In the longitudinal study, students from Grade 3 and 4 showed significant increase in place value understanding as described by the model over the course of 1 year. A cross-sectional study in a different cultural and educational environment (South Africa) provides similar empirical evidence in support of the model validity (Herzog et al., 2017).

Pre-decadic Level: Initially, children perceive multi-digit numbers as entities without any decimal structure (Cobb and Wheatley, 1988; Ross, 1989; Fuson et al., 1997a). Children might be able to decompose numbers in general (e.g., 24 into 12 and 12). However, canonical decompositions into tens and units (e.g., 2 tens and 4 units) have no specific base-ten related quality to these children. The canonical decomposition as fundamental construction principle of numbers within the decimal place value system is only one of many possible decompositions and children are unlikely to recognize tens and units in this decomposition at the Pre-decadic Level.

Place Values (Level I): At first, children understand that the digits in numbers can be mapped to decimal bundling units such as units, tens, etc. (Cobb and Wheatley, 1988; Ross, 1989; Fuson, et al., 1997b). Based on this place value understanding, children at this level can map digits to corresponding bundling units. However, they do not yet understand, how the bundling units are related. That means that children at Level I can handle canonical decompositions (e.g., 2 tens and 4 units) based on the place value understanding of this level, but struggle with so-called non-canonical decompositions (e.g., 1 ten and 14 units).

Tens-Units Relation with Visual Support (Level II): At Level II, children develop an understanding of the relation of tens and units that is based on visual support (Cobb and Wheatley, 1988; Steffe, 1992). They can bundle and unbundle tens and units, but rely on counting processes to verify the equivalence of ten units and one ten. It takes decimally structured material such as base-ten blocks for them to reliably employ these counting processes (Fuson, et al., 1997a; Nührenbörger and Steinbring, 2008). Non-canonical representations that are given abstractly cannot be handled based on this place value understanding. Therefore, children might identify "ten" rather with the corresponding visualization (e.g., a tens stick, Van de Walle et al., 2016) than with a composition of ten units. Children at this level understand the relation of the bundling units only for units and tens, while bigger bundling units (e.g., hundreds) are not yet integrated (Scherer and Moser Opitz, 2010).

Tens-Units Relations without Visual Support (Level III): Level III is characterized by two developmental changes in place value understanding. First, children detach from visual representations for two-digit numbers which enables them to handle non-canonical representations of tens and units. Based on an interiorized understanding of the tens-units relation, they do not necessarily need counting routines to verify that ten units make up one ten. Second, children extend the representation-based understanding of the relation between the bundling units onto bigger units such as hundreds, thousands, etc. Like on Level II for two-digit numbers, children need decimally structured material to employ the iterative bundling principle for multi-digit numbers in general. Scherer and Moser Opitz, (2010) emphasize the necessity of bundling tens to hundreds to understand the iterative bundling principle.

General Bundling Units Relations (Level IV): At the fourth and final level of the model, children have successfully established an abstract concept of the relation between the bundling units. Similar to the transition from Level II to Level III for the relation between tens and units, children now detach from concrete representations for multi-digit numbers. As children established a profound understanding of the positional principle as well as the iterative bundling principle with this learning step, their place value understanding development is considered completed.

The Current Study

According to the ADAPT model, the rules employed during transcoding are asemantic, procedural, and developmental. The asemantic characteristic means that transcoding processes are not necessarily bound to the numerical meaning of the number words. The procedural characteristic refers to the transcoding rules. As children follow algorithms during transcoding processes, they may not actually understand the structure behind the processes. The developmental characteristic highlights that children actively develop the numerical lexicon for number words. With increasing experience, children gain routine in employing the transcoding rules. Thus, their increasing transcoding proficiency in primary school is an outcome of learning processes (Barrouillet et al., 2004; Moura et al., 2015).

Research mostly focused on working memory and phonological awareness as influencing factors on transcoding. However, given the interrelation of procedural and conceptual knowledge, the assumptions of the ADAPT model also rise questions regarding the conceptual fundaments of transcoding processes (Rittle-Johnson and Schneider, 2015). As the process of transcoding itself is asemantic, the numerical meaning of a number is not activated during transcoding. However, there has to be some kind of meaning that numbers have to children. It is unclear, to what extent children's understanding of numbers and decimal bundling units–although not activated during transcoding–are related to their transcoding skills. Transcoding is also procedural, which means that the transcoding rules are employed by routine. Thus, the ADAPT model makes no predictions on the influence of children's conceptual understanding on transcoding. As procedural and conceptual knowledge are interrelated, we cannot exclude an influence of conceptual place value understanding on transcoding. Finally, the ADAPT model describes transcoding as developmental. This implies that children gain knowledge that is relevant for transcoding during primary school. What knowledge is relevant for transcoding, and in particular the role of conceptual knowledge in this regard, is not finally investigated. To sum up, this study aims at investigating the role of conceptual place value understanding for transcoding skills to better understand the cognitive prerequisites of transcoding processes.

To address these questions at least partially, we investigated transcoding abilities in relation with place value understanding as described by the Herzog et al. (2019) model in German second and third graders. As place value understanding usually develops in Grades 3 to 5 (Herzog et al., 2019; Houdement and Tempier, 2019), and transcoding abilities substantially improve during Grades 1 to 3 (Moura et al., 2015), children in Grade 2 and 3 are of particular interest regarding this research question. The model allows assessing and localizing children's individual status of place value understanding within the developmental level sequence. This approach facilitates to investigate differences in transcoding performance between children with more or less elaborated place value understanding. As the model provides qualitative descriptions of children's place value understanding, differences in transcoding abilities might not only be explained in terms of task performance, but also in terms of children's ideas of the place value system. The qualitative description of children's place value understanding based on the model by Herzog et al. (2019) in relation to transcoding abilities might give insights into the conceptual underpinnings of transcoding.

Besides variance in error rates across children at different levels of place value understanding, we expect variance in error types to provide substantial information on conceptual underpinnings of transcoding. At least two aspects are highlighted in research. First, lexical and syntactical errors might differ in the way they are related to place value understanding. As lexical errors are mostly mapping errors between digits and numbers, they might be less related to place value understanding than syntactical errors (Zuber et al., 2009; 215; Moura et al., 2013). Second, the German sample allows investigating inversion related errors. Regarding the cognitive foundation of inversion errors, two competing approaches can be found in the literature. While mathematics education research considers inversion related errors as an indicator for low place value understanding (e.g., Schulz, 2014), (neuro-) psychological studies stress the influence of working memory on inversion related errors (Bahnmueller et al., 2015; Pixner et al., 2016). This study aims at contributing to this debate by investigating the influence of place value understanding and inversion related errors.

Two main research questions (RQ) will structure the investigation of cognitive underpinnings of transcoding of this study. The qualitative level description of the model of place value understanding allows making testable predictions regarding the differences in transcoding performance between children at different levels.

RQ1: To what extent does transcoding performance vary between children at different levels of place value understanding as described by the model by Herzog et al. (2019)? We expect that children at higher levels of place value do fewer transcoding errors in general. More specifically, the concept described in Level I (identification of place values) is supposed to support transcoding processes (Bahnmueller et al., 2018; Herzog et al., 2019). Thus, especially children at the Pre-decadic Level are expected to show lower transcoding performance, while children at higher levels are expected to differ only slightly regarding transcoding performance.

RQ2: How are different transcoding error categories such as lexical and syntactical as well as inversion related errors interrelated with place value understanding? Based on the

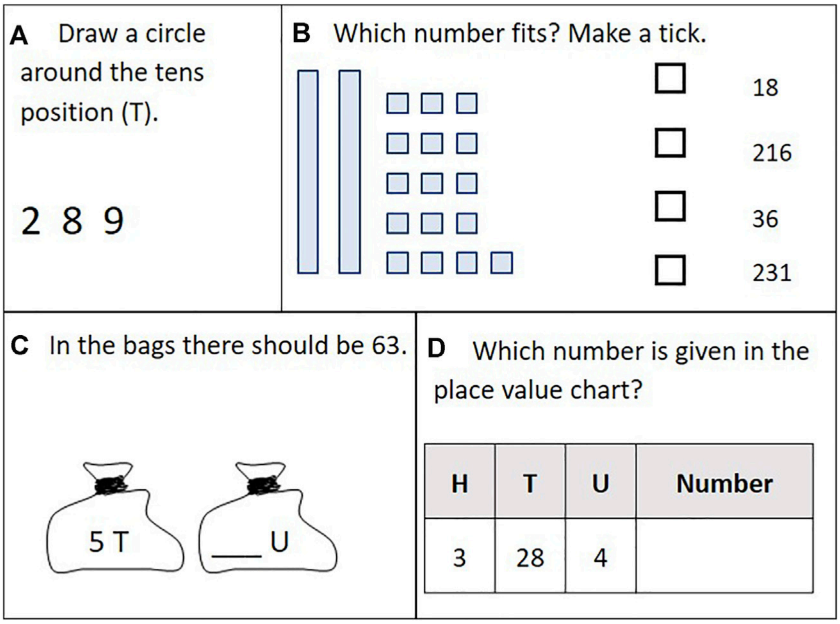


FIGURE 1 | Example items from the place value understanding assessment for Levels (I–IV)**(A–D)**.

underlying processes in transcoding as described in the ADAPT model, we hypothesize that syntactical errors are negatively associated with place value understanding, while lexical errors may not or only slightly be associated with place value understanding. In particular children, who do not have a conceptual basis for the identification of place values (Pre-decadic Level) are expected to make syntactical errors. The study design allows to investigate the effect of place value understanding on inversion related errors, which might contribute to the debate on the cognitive foundation of inversion-related errors.

MATERIALS AND METHODS

Sample

In total 266 students participated in the study. Of the total sample, 135 students (69 female, M_{age} = 91.4 months, SD_{age} = 5.7 months) were in Grade 2 and 131 (57 female, M_{age} = 104.3 months, SD_{age} = 6.4 months) were in Grade 3. In both grades, data was collected during the first 3 months of the school year. Children were acquired from three schools of which one was located in an upper-class, one in a middle-class, and one in a lower-middle-class suburb. Written consent to participate in the study was acquired in advance from the parents. The study was approved by the local ethics committee of the authors' university.

The sample was selected based on the contents of the mathematics curriculum. In Grade 1, mathematics classes cover the number range up to 20, in Grade 2 up to 100, and in Grade 3 up to 1,000. Thus, children, who just entered Grade 2 or Grade 3 are appropriate for the aims of this study. As no inclusion criteria were applied, all second- and third-graders from the selected schools participated in the study, if consent was obtained.

Instruments

Transcoding

Transcoding performance was assessed by a writing-numbers-test. Children were given 28 single-to four-digit numbers verbally. The stimuli were identical with Moura et al. (2015) and can be found in **Figure 1**. Transcoding errors were categorized similar to Moura et al. (2015). Incorrect mappings of numbers and digits (e.g., 961 for 951) were coded as lexical errors. Errors violating the procedural rules (e.g., 90051 for 951) were coded as syntactical errors. As a specific type of syntactical errors, inverted tens and units (e.g., 915 for 951) were coded as inversion-related errors. Based on the rules of the ADAPT model involved in the transcoding process, the numbers were categorized in numbers with null, low, medium, and high complexity (Moura et al., 2015).

Place value Understanding

Based on the model by Herzog et al. (2019), two item collections were used to assess children's place value understanding. Because children in Grade 2 are not yet introduced to numbers bigger than hundred and nearly no child had developed the concept of Level IV in a piloting study, we omitted Level IV items in Grade 2. The items for both Grades were based on the item collection which had been used in another study to validate the model in Germany. A one-dimensional Rasch analysis confirmed that the item difficulties were coherent with the assumptions of the theoretical model (Herzog and Fritz, 2019). Example items from the place value understanding assessment are presented in **Figure 1**. The full item collections are available by request to the corresponding author.

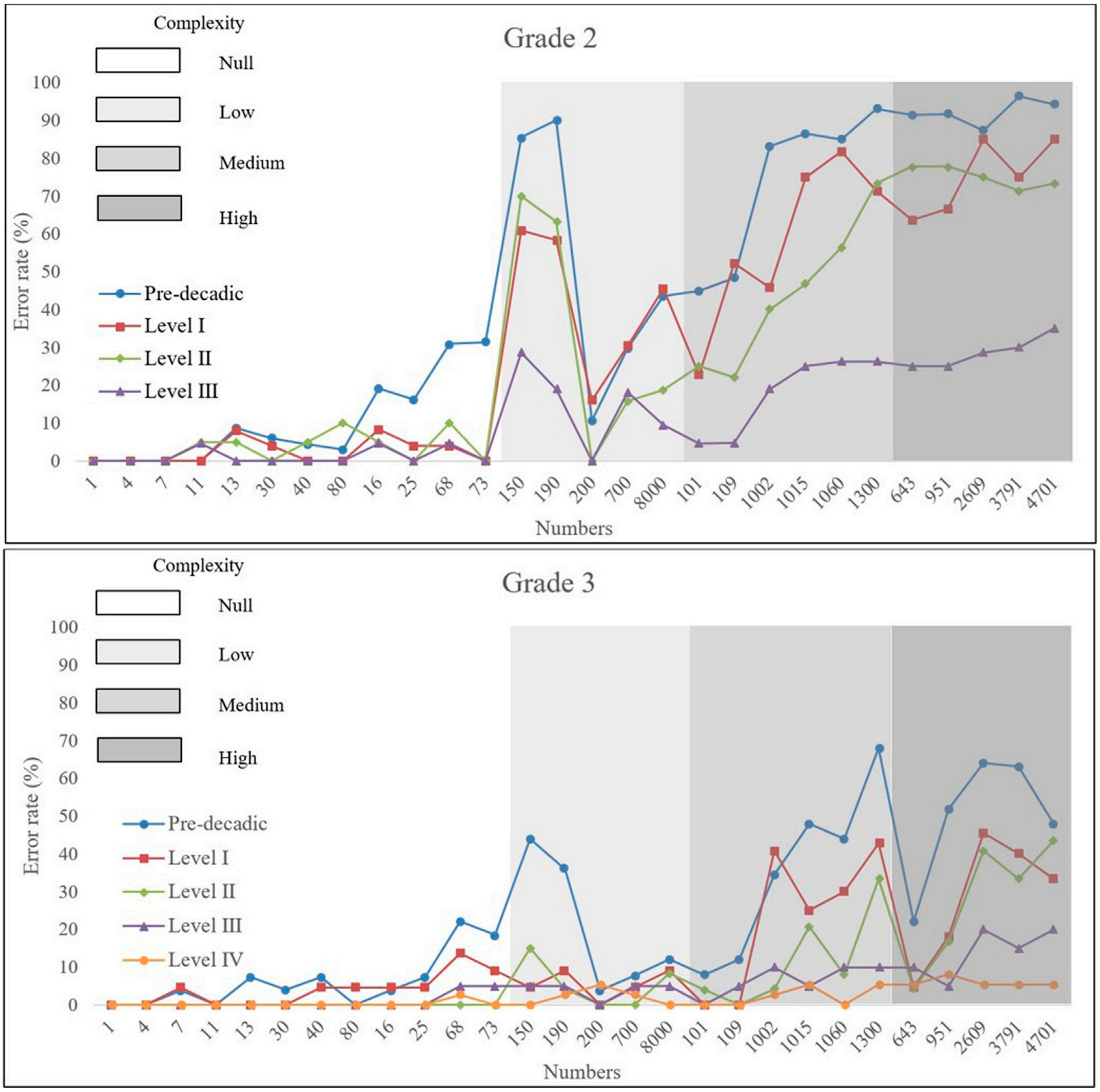


FIGURE 2 | Error rates for children at different levels of place value understanding in Grade 2 and Grade 3 for the stimuli of the transcoding tasks. Numbers ordered by level of complexity (background shades).

In Grade 2, 40 items aligned to the Levels I to III were employed. 16 items were aligned to Level I, while 12 items corresponded to the Levels II and III each. As items were presented in random order, more items on the first level were included in the assessment to prevent frustration and tiring in the children. 20 items (Level I: 6, Level II: 8, Level III: 6) were original items of the validation study and all additional items were variations of these items. The internal consistency of the 40 items was good (Crombach's α =0.934).

The item collection for Grade 3 contained 48 items. Of the 48 items, 16 were aligned to Level I, 12 items were aligned to Level II and III, and 8 items were aligned to Level IV. Lower levels were overrepresented for the same reasons as in Grade 2.30 items (Level I: 7, Level II: 8, Level III: 7, Level IV: 8) were identical with the version of the validation study. The additional items on Levels I to III were identical to those included in the assessments for Grade 2. Internal consistency for the Grade 3 item collection was good (Crombach's α =0.947).

Children's individual level of place value development was assessed based on percentage of correct answers per level. If children solved at least 75% of the items of a level, they were assigned to the corresponding level. The highest achieved level was recorded as the current conceptual level of place value understanding (for similar approaches see Lee and Sarnecka, 2009; Ricken et al., 2013; Fritz et al., 2017; Balt et al., 2020). To address lucky guesses and mistakes, children did not have to solve all items of a level to be considered having developed the concept of the level. Determining individual level achievement by a

TABLE 1 | Descriptive analysis of the place value understanding.

Level	Grade 2		Grade 3	
	Level accomplishment	PVU (max. 40)	Level accomplishment	PVU (max. 48)
	n (%)	*M (SD)*	n (%)	*M (SD)*
Pre-decadic	68 (50.4)	12.31_a (5.52)	27 (20.6)	18.74_a (5.16)
Level I	25 (18.5)	22.60_b (4.99)	22 (16.8)	25.18_b (5.47)
Level II	20 (14.8)	26.55_b (5.16)	25 (19.1)	30.64_c (5.08)
Level III	22 (16.3)	34.95_c (2.68)	20 (15.3)	36.45_d (3.00)
Level IV			37 (28.2)	40.14_d (6.06)
Total	135	20.01 (9.95)	131	30.84 (9.58)

**Note. PVU, raw scores of the place value understanding tests; M, mean, SD, standard deviation. Means sharing the same subscript do not differ.*

criterion of 75% also visualizes the assumption of the levels as overlapping waves, as the development of place value understanding is not considered disruptive (i.e., replacing prior knowledge at once), but as a progressive elaboration (Siegler and Alibali, 2005; Clements and Sarama, 2014). The internal hierarchy of the levels showed in this analysis, too. In all cases, children also fulfilled the 75% criterion for lower levels than the achieved level. For example, a child that solved 75% of the items at Level III also met this benchmark for Levels I and II.

Data Collection

The data collection was conducted in the classrooms during usual lesson hours. Teachers were informed and received place value training material as an incentive for participation. Two trained undergraduate students helped with the data collection. Both received an intensive training on data collection beforehand and were supervised by the authors.

Besides the presented instruments, number line estimation and early arithmetic concepts were also assessed (but not included in this study). To minimize cognitive load, the tests were split into two sections. The tests analyzed in this study—transcoding and place value understanding—were in the same section and thus assessed on the same day. In all assessments, transcoding was assessed first, as children needed to write down numbers synchronously. The item order in the transcoding test was aligned to Moura et al. (2013; see also **Figure 2**). Place value understanding was assessed subsequently, because children could solve these tasks individually and in their own speed. All children had enough time to finish the place value assessment. The order of the items in the place value test were randomized across the levels. Thus, easier and more difficult items alternated. The items from the place value assessment were arranged into two sets to avoid position effects.

RESULTS

In Grade 2, half of the children were assessed at the Pre-decadic Level. The other half of the second-graders and all third-graders were nearly equally distributed across the levels of place value understanding. One-factorial ANOVAs validated the differences in place value performance between the children at different place value levels in Grade 2 ($F(3, 131) = 132.064$, $p<.001$, $\eta^2 = 0.752$) and Grade 3 ($F(4, 126) = 77.839$, $p<.001$, $\eta^2 = 0.712$). With two exceptions—Level I vs. II in Grade 2 ($p = 0.058$) and Level III vs. IV in Grade 3 (p=.112)—Bonferroni-corrected post-hoc tests showed significant differences in place value understanding between the level-subgroups in both grades. Thus, the classification of the participating children appears appropriate. The distribution of children to the levels and the corresponding descriptive statistics are summarized in **Table 1**.

As expected, transcoding performance was positively associated with place value understanding in both grades (RQ1). Raw scores of correct answers in transcoding tasks and place value understanding tasks correlated substantially in Grade 2 ($r = 0.638$, $p < 0.001$) and Grade 3 (r=.585, $p < 0.001$). **Table 2** shows mean correct answers in the transcoding test for children at different levels. A one-way ANOVA confirmed that the group differences were significant in Grade 2 ($F(3, 130) = 25.223$, $p<.001$, $\eta^2 = 0.368$) and Grade 3 ($F(4, 125) = 12.476$, $p < 0.001$, $\eta^2 = 0.285$). Post-hoc tests (Bonferroni) revealed consistent significant lower performance only in children at the Pre-decadic Level ($p < 0.019$ for all comparisons).

An analysis of the performance regarding numbers of different complexity provides deeper insights in the associations of transcoding and place value understanding. Transcoding complexity of stimuli was determined by the number of procedural rules in terms of the ADAPT model that have to be applied to the particular number as described by Moura et al. (2015). Regarding the numbers with the lowest complexity (Null), there were no significant differences between children at different levels (I to III or IV) of place value understanding. We found ceiling effects in all subgroups of place value understanding for the least complex numbers. With increasing number complexity, differences between the children with different place value understanding got bigger in Grade 2. In Grade 3, transcoding performance in children at the Pre-decadic Level dropped with increasing complexity of numbers. However, ceiling effects were found in children at Levels I to IV for numbers of null, low, and medium complexity. Children at Levels III and IV showed ceiling effects even for the most complex numbers. In summary, the effect of place value understanding on transcoding applied primarily for complex numbers and children at the Pre-decadic Level.

TABLE 2 | Means and standard deviations of transcoding performance in children at different levels in Grade 2 and 3.

Level	Grade 2					Grade 3				
	Total (max. 28)	Task complexity				Total (max. 28)	Task complexity			
		Null (max. 12)	Low (max. 5)	Medium (max. 6)	High (max. 5)		Null (max. 12)	Low (max. 5)	Medium (max. 6)	High (max. 5)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Pre-decadic	14.65_{a} (3.49)	10.74_{a} (1.74)	2.30_{a} (1.17)	1.46_{a} (1.48)	0.36_{a} (0.95)	20.59_{a} (5.83)	11.07_{a} (1.71)	3.78_{a} (1.12)	3.73_{a} (1.66)	2.37_{a} (1.88)
Level I	17.72_{b} (4.69)	11.68_{b} (0.75)	$2.72_{a,b}$ (1.54)	$2.24_{a,b}$ (1.83)	1.08_{a} (1.66)	24.27_{b} (3.76)	11.55_{a} (1.37)	4.73_{b} (0.63)	$4.45_{a,b}$ (1.63)	$3.50_{a,b}$ (1.63)
Level II	18.55_{b} (5.11)	$11.60_{a,b}$ (0.60)	$3.10_{a,b,c}$ (1.41)	2.85_{b} (2.21)	1.05_{a} (1.87)	$24.76_{b,c}$ (4.94)	11.64_{a} (1.80)	4.56_{b} (1.04)	5.16_{b} (1.28)	$3.54_{a,b,c}$ (1.56)
Level III	23.86_{c} (5.46)	11.82_{b} (0.40)	$4.14_{b,c}$ (1.39)	4.59_{c} (1.99)	3.43_{b} (2.11)	$26.65_{b,c}$ (2.85)	11.90_{a} (0.45)	4.80_{b} (0.62)	5.60_{b} (1.19)	$4.30_{b,c,d}$ (1.38)
Level IV						27.47_{c} (1.75)	11.97_{a} (0.16)	4.89_{b} (0.39)	5.86_{b} (0.54)	4.70_{d} (0.91)

Note. M, mean; SD, standard deviation. Means sharing the same subscript do not differ.

TABLE 3 | Means and standard deviations of lexical, syntactical and inversion related errors in children at different levels in Grade 2 and 3.

Level	Grade 2			Grade 3		
	Lexical	Syntactical	Inversion related	Lexical	Syntactical	Inversion related
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Pre-decadic	2.06_{a} (1.98)	9.10_{a} (4.32)	1.13_{a} (1.42)	1.96_{a} (3.71)	4.19_{a} (4.05)	0.30_{a} (0.61)
Level I	$1.40_{a,b}$ (1.73)	$7.00_{a,b}$ (4.65)	0.32_{b} (0.56)	$0.55_{a,b}$ (0.96)	$2.82_{a,b}$ (3.16)	0.18_{a} (0.50)
Level II	$1.20_{a,b}$ (1.06)	$6.00_{b,c}$ (3.88)	$0.60_{a,b}$ (0.88)	0.36_{b} (0.57)	$1.92_{b,c}$ (2.60)	0.00_{a} (0.00)
Level III	0.59_{a} (1.01)	2.64_{c} (4.16)	0.32_{b} (0.65)	0.05_{b} (0.22)	1.40_{c} (2.84)	0.15_{a} (0.49)
Level IV				0.22_{b} (0.53)	0.35_{c} (1.36)	0.03_{a} (0.16)

Note. M, mean, SD, standard deviation. Means sharing the same subscript do not differ.

In line with Moura et al. (2015), children in Grade 3 performed better in the transcoding tasks than the second graders, even when they were at the same level of place value understanding. As for place value understanding, the effect of the Grade increases with number complexity. Within the Grade-specific subgroups, no significant correlation of chronological age and transcoding performance was found (Grade 2: $r = -0.071$, $p = 421$; Grade 3: $r = -0.055$, $p = 0.533$). We therefore conclude that, besides place value understanding, transcoding is influenced by schooling, as children in Grade 3 have been introduced to three-digit numbers.

A comparison of the error rates regarding the 28 numbers used in the transcoding tasks across children at different levels of place value understanding (**Figure 2**) with a comparable analysis regarding children in different grades (Moura et al., 2015) reveals similar patterns. Different Grade-levels and levels of place value understanding have nearly identical effects on the error rates. The similarities can even be found regarding the individual stimuli used in the transcoding tasks.

In line with previous research, children made substantially more syntactical errors than lexical errors (Moura et al., 2015). An analysis of the error types revealed the same significant effects of grade and place value understanding for all error types (RQ2). As comprised in **Table 3**, individual error rates decreased with increasing level of place value understanding, and children in Grade 3 made less errors than second graders. However, these effects did not apply to all error types to the same extent. The individual rates of syntactical errors were stronger affected by place value understanding and Grade level than lexical error rates. Different sensitivity to place value understanding is visualized by effect sizes of one-way ANOVAs, which were higher for syntactical errors (Grade 2: $\eta^2 = 0.233$, Grade 3: $\eta^2 = 0.198$) than for lexical errors (Grade 2: $\eta^2 = 0.097$, Grade 3: $\eta^2 = 0.137$).

Inversion related errors are a specific subtype of syntactical errors found in German. Pure inversion related errors only made up a small percentage (Grade 2: 10.7%, Grade 3: 6.1%) of the syntactical errors. Most syntactical errors were not inversion related, but were related to incorrect integration of hundreds (e.g., 90051) or thousands (e.g., 10002 for 1002). As for lexical errors, the effect of place value understanding on inversion related errors was rather low (Grade 2: $\eta^2 = 0.103$, Grade 3: $\eta^2 = 0.073$). This finding indicates that inversion related transcoding errors are only partially dependent on missing place value understanding. In contrast to place value understanding, Grade level had a bigger effect on inversion related errors, as the rate for inversion related errors in Grade 3 was only one sixth of the rate in Grade 2 for the whole sample. Only in children at the Pre-decadic Level there was no effect of the grade level.

DISCUSSION

The results of this study show that transcoding abilities are associated with place value understanding in general (RQ1). This association was slightly stronger for second graders than for third graders and for more complex numbers. However, the differences between second- and third-graders were not significant. Especially children who have not yet developed a conceptual basis for identifying place values (Pre-decadic Level) showed lower transcoding performance than children at higher levels. The association between place value understanding and transcoding did not apply to all types of errors in the same way (RQ2). The biggest effect of place value understanding was found regarding syntactical errors. In contrast to findings reported in the literature, inversion related errors occurred only rarely (Zuber et al., 2009; van der Ven et al., 2017). As well as inversion related errors, lexical errors were only little affected by differences in place value understanding. Children at the Pre-decadic Level showed a smaller difference in error rates between Grade 2 and Grade 3. At the same time, these children performed significantly lower in the transcoding tasks than children at higher levels. We suspect that this group might be most likely to develop persistent difficulties with transcoding and even mathematical learning difficulties.

Cognitive Foundations of Transcoding

Independently from the level of place value understanding, children in Grade 3 showed higher transcoding proficiency than children in Grade 2. Thus, place value understanding may explain transcoding performance only partially. Experience with writing numbers due to exposure and formal instruction in school have to be considered another basis for children's development of transcoding abilities, too. As age did not correlate with transcoding proficiency, transcoding abilities appear not as a result of maturing, but rather a result of education. These findings indicate that transcoding abilities are supported by, but not completely bound to an elaborated place value understanding.

Based on the data gained in this study, we can only speculate which learning contents of formal schooling promote transcoding during Grades 2 and 3. Informed by the curriculum, decimal arithmetic strategies such as break-down (27 + 15 = 20 + 10 + 7 + 5 = 30 + 12 = 42) or decompositions (27 + 15 = 27 + 3+ 12 = 30 + 12 = 42) seem reasonable contents that support children's transcoding abilities. It is also likely that children gain experience during primary school when asked to write, read, and compare numbers. The better transcoding performance regarding more complex numbers in third graders supports this suggestion, as they have been introduced to three-digit numbers in school. More insight in schooling effects might be obtained by targeted intervention studies in Grade 2 in which the different contents can be taught separately. In addition, longitudinal single case studies over a longer time period can align the development of transcoding abilities to the development of place value understanding on the one hand and formal schooling contents on the other hand.

The findings of this study regarding the error types can be brought in line with the ADAPT model. First, the association of place value understanding and transcoding was dominantly caused by syntactical errors. Syntactical errors occur when children incorrectly apply the procedural rules of the ADAPT model. Knowledge of these rules partially corresponds with place value understanding as described by the Herzog et al. (2019) model. For instance, the knowledge of the place values (Level I) can be identified as conceptual underpinning of the P2/P3-rules. Second, lexical errors were less associated with place value understanding. Lexical errors are supposed to stem from incorrect mappings of number words and Arabic symbol. Mapping errors—may they be due to falsely learned number words or mistakes when deriving from the long-term memory—are based rather on access difficulties to the numerical lexicon than on the procedural rules described in the ADAPT model. Third, a possible reason for inversion errors highlighted in research are working memory capacities (Zuber et al., 2009): When presented an inverted two-digit number word (e.g., five-and-twenty [25 in German]), children have to keep the units part (five) in mind while waiting for and writing down the tens part (twenty). Thus, the ten-unit-inversion draws on the verbal working memory and the central executive (storage of the unit), inhibition (not writing the unit) and visual-spatial working memory (coordinating left and right when writing down the digits). The small association of place value understanding and inversion errors suggests that inversion errors are mostly bound to working memory, but only slightly on place value understanding (Bahnmueller et al., 2015; Pixner et al., 2016). Further studies in which the influence of place value understanding and working memory can be compared directly, for example in a multiple regression model, are needed in this regard.

In both grades and all levels of place value understanding, more complex numbers had higher error rates. However, the effect of number complexity did not apply to all numbers in the same way. For instance, the numbers 200, 700, and 8000 (so-called X00 and X000 numbers; Zuber et al., 2009) were not affected by the number complexity effect. These number words are directly composed by a single digit number word (e.g., seven) and the decimal bundling unit word (e.g., hundred). Obviously, children struggled less with X00 and X000 numbers than with other numbers in the same number range and the same level of complexity. This is especially interesting, as error rates rise very fast for the first numbers of low complexity (150 and 190, so-called XX0 numbers; Zuber et al., 2009). It seems as if X00 and X000 number words are more intuitive to children. The unexpected low error rate on these numbers even in children at the Pre-decadic Level—who do not have any understanding of the decimal place value system yet—underpins this explanation. In terms of the ADAPT model, the different error rates for XX0 numbers and X00/X000 numbers partially contradict the assumption that a number like 700 is transcoded by applying the rules P1 (seven: fact retrieval), P2 (hundred: open three-digit frame), and P4 (end of number word: fill in the empty slots with zeros). Given the error rates in this study, X00/X000 number words are likely to be retrieved from long-term memory just as one- and two-digit numbers in children in Grade 2 and 3. Comparable findings from Moura et al. (2015) support this hypothesis.

Transcoding is a Part of Place Value Understanding

While children at the lower levels of place value understanding performed lower in the transcoding tasks in general, 32.4% of the third graders up to Level II solved all transcoding tasks correctly. Although lacking a profound conceptual understanding of the place value system, these children did not struggle even with the most complex transcoding tasks. In Grade 3, children up to Level II showed ceiling effects except for the most complex numbers. Moura et al. (2015) reported that the vast majority of the children at the end of Grade 4 can perform transcoding tasks even when numbers are complex, while Herzog et al. (2019) showed that about half of the children at this age have not yet developed the concepts of Level III and IV. Obviously, a profound understanding of the place value system includes transcoding abilities. But against the background of the findings listed up above, we suggest that place value understanding covers more knowledge than transcoding alone. Especially an automatized understanding of the relation of the bundling units (Level III and IV) goes beyond transcoding.

Place value understanding can be differentiated into procedural and conceptual aspects (Rittle-Johnson and Schneider, 2015; Van de Walle et al., 2016; Herzog et al., 2019): While procedural place value understanding refers to the knowledge how the elements of the place value system interact, conceptual place value understanding covers the knowledge why they interact. Following the models of transcoding and place value understanding that formed the basis of this study, transcoding has to be considered a rather procedural aspect of place value understanding. This notion of transcoding is in line with the ADAPT model being a procedural model. But procedural aspects usually have a conceptual underpinning, too. Knowledge of the bundling units (Level I) appears as a reasonable conceptual basis for transcoding by theoretical considerations and the empirical results of this study. This idea was also the hypothesis for both research questions. The results of this study support this assumption partially. Children at the Pre-decadic Level performed significantly lower in the transcoding tasks than children at higher levels. At the same time, transcoding performance increased with higher level of place value understanding. Comparable group differences were found in regard to the error types. Thus, knowledge of Level I coincides with better transcoding performance as hypothesized, but the association of place value understanding with transcoding performance is not limited to the knowledge of Level I.

Research has given examples, how children can solve tasks procedurally without having a profound conceptual foundation (Selter, 2001). This might explain that 18.5% of the third graders at the Pre-decadic Level could solve all transcoding tasks in this study. These results of the study underpin the notion of transcoding as a rather procedural aspect of the model of place value understanding. An analysis of procedural and conceptual aspects of place value understanding supports the claim that place value understanding covers skills and knowledge beyond transcoding, while transcoding is an integral part but not totally incremented in place value understanding (Herzog et al., 2019).

As transcoding is a basal component of place value understanding, difficulties with transcoding tasks may be an easy to use screening tool to identify children who struggle severely with developing a profound place value understanding (Moura et al., 2015). Especially complex numbers can be helpful to find children at-risk for difficulties with place value understanding. Keeping in mind the effect of formal schooling, transcoding seems to be most useful in early primary school. Given the low prevalence of children who lack transcoding abilities, transcoding tasks may only identify children with severe difficulties within the range of mathematical learning difficulties. However, transcoding appears not as a useful screening for children with mild math difficulties which are limited to distinct aspects of mathematics.

The Role of Place Value Relations in Place Value Processing

The proposed notion of place value understanding and transcoding as overlapping domains of knowledge that are neither included nor disjunct has implications for the taxonomy of place value processing (Nuerk et al., 2015). If, as the results of this study suggest, transcoding conceptually relies both on place value identification (Level I) and an understanding of the relations between the bundling units (Levels II-IV), the latter might be added to the Nuerk et al. (2015) taxonomy. *Place value relations* could specify the conceptually grounded knowledge, how decimal bundling units are related and how they can be traded. This level of place value processing would especially cover the handling of non-canonical number representations with and without visualizations.

Place value relations can be distinguished from the three other levels (place value identification, place value activation, and place value computation). Place value identification can be identified with Level I (Nuerk et al., 2015; Herzog et al., 2019). The hierarchy of the developmental model of place value understanding shows that place value relations exceed place value identification (Herzog and Fritz, 2019). Place value activation refers to the numerical value of the positions in multi-digit numbers. The automatic activation of the numerical information of place values gets visible for example in number comparison (Nuerk et al., 2015). Nuerk et al., (2001) have shown that numerical information is activated separately for tens and units. The compatibility effect in number comparison, which leads to higher reaction times in comparing incongruent number pairs (e.g., 72 and 58) than congruent pairs (e.g., 72 and 61) visualizes this automatic information: The (unnecessary) information regarding the units is activated automatically and independently from the information of the tens. As place value activation affects the decimal bundling units separately, it is unlikely that this level of place value processing can be

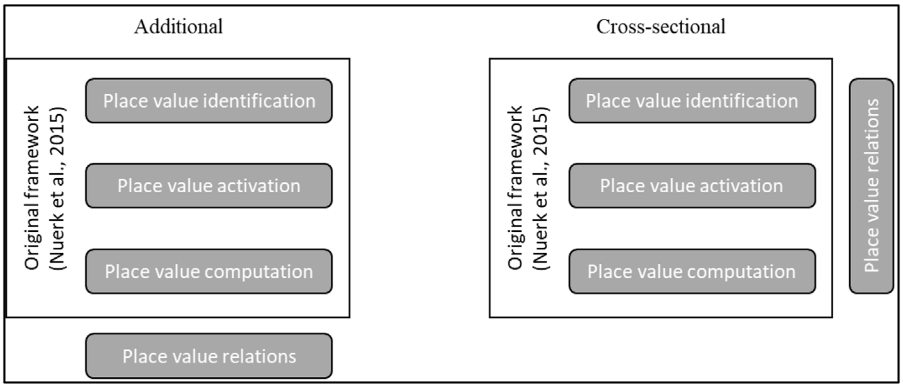


FIGURE 3 | Possible conceptualizations of place value relations within the framework of levels of place value processing (Nuerk et al., 2015).

identified with place value relations. Place value computation specifies the processing of place value information when solving arithmetic tasks. Studies have shown that tasks involving carries (e.g., 27 + 18) are more difficult than tasks without carries (Imbo et al., 2007; Klein et al., 2010; Moeller et al., 2011b). In particular children with mathematical learning difficulties struggle with tasks involving carries (Lambert and Moeller, 2019). As carry tasks require to coordinate tens and units while computing, one might suspect that place value computation could be identified with place value relations. However, results from the validation studies in Germany and South Africa of the Herzog et al. (2019) model show that solving addition and subtraction tasks with carries does not necessarily imply a profound and abstract understanding of the relation of the bundling units (Herzog et al., 2017; Herzog and Fritz, 2019).

While place value relations as described here are not precisely covered by the existing levels of place value identification, there are possible interactions. As underpinned by the model hierarchy, place value identification (Level I) is a basis for place value relations. Place value computation correlates with general mathematical skills which include place value relations (Lambert and Moeller, 2019; Herzog and Fritz, 2020). These findings rise the question, how place value relations can be conceptualized within the framework of place value processing (Nuerk et al., 2015). At least two conceptualizations seem reasonable (see **Figure 3**): First, place value relations might be an additional fourth level of number processing. In this case, variance in place value relations would lead to different proficiency for example when handling non-canonical number representations, but not interact with effects associated with place value identification, activation, or computation. Second, place value relations could be a cross-sectional level that interacts with place value identification, activation, and computation to some extent. Such a conceptualization would imply that place value relations cannot completely be disentangled from the other levels of place value processing. Empirical research might provide evidence to decide which conceptualization is more accurate.

Limitations and Future Perspectives

At least three limitations of this study have to be considered. Each of them might inform future research on the association of place value understanding and transcoding. First, this study included no cognitive or mathematical control variables. Given the body of research on the relation between transcoding and working memory, the above made hypotheses on error types and their varying foundation on place value understanding on the one hand and working memory on the other hand deserve more detailed investigation. While transcoding and place value understanding individually predict later arithmetic performance (Moeller et al., 2011a; Herzog and Fritz, 2020), their interaction remains unclear. Based on the claim that place value understanding covers and exceeds transcoding, one might speculate that place value understanding mediates the influence of transcoding on arithmetic performance to some extent.

Second, the ADAPT model and the Herzog et al. (2019) model are supposed to be developmental. That means that children show substantial progress over time in the skills described in these models (e.g., Herzog and Fritz, 2019). The interaction of developmental trajectories in place value understanding and transcoding skills can only be evaluated validly in a longitudinal study. Results from such a longitudinal study could provide valuable insights in the development of transcoding abilities. While the ADAPT model gives a clear description on the cognitive processes in transcoding, the trajectories children follow while learning how to apply these processes remain vague. For example, the structure of the Herzog et al. (2019) model suggests that P2 (two-digit numbers) and P3 (three-digit numbers) rules are learned successively. A developmental progression of transcoding that could be derived from a longitudinal study could answer this question. Additionally, such a developmental progression would help to structure transcoding instruction for children struggling with writing numbers (Clements and Sarama, 2004).

Third, in this study transcoding was only operationalized in form of writing numbers. Against the theoretical background of the ADAPT model (reading numbers) and the Triple Code Model (magnitude representations), substantial transcoding paths have not yet been investigated regarding their association with place value understanding. For reading numbers, similar effects as found in this study seem likely, as transcoding processes specified in the ADAPT model hold for both directions.

Predictions on the conceptual basis of magnitude representations such as number line estimation appear more difficult and require further research.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article are available by the authors on request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local ethics committee of the University of Wuppertal. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

Study design: MH and AF; Data collection: MH; Analyses: MH; Manuscript edition: MH and AF.

FUNDING

The publication of this article was funded by the publication fonds of the University of Wuppertal (funding number: GoldOA-PUBL20201116).

REFERENCES

Bahnmueller, J., Moeller, K., Mann, A., and Nuerk, H. C. (2015). On the Limits of Language Influences on Numerical Cognition - No Inversion Effects in Three-Digit Number Magnitude Processing in Adults. *Front. Psychol.* 6, 1216. doi:10.3389/fpsyg.2015.01216

Bahnmueller, J., Nuerk, H. C., and Moeller, K. (2018). A Taxonomy Proposal for Types of Interactions of Language and Place-Value Processing in Multi-Digit Numbers. *Front. Psychol.* 9, 1024. doi:10.3389/fpsyg.2018.01024

Balt, M., Fritz, A., and Ehlert, A. (2020). Insights into First Grade Students' Development of Conceptual Numerical Understanding as Drawn from Progression-Based Assessments. *Front. Educ.* 5, 80. doi:10.3389/feduc.2020.00080

Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: A Developmental, Asemantic, and Procedural Model for Transcoding from Verbal to Arabic Numerals. *Psychol. Rev.* 111 (2), 368–394. doi:10.1037/0033-295X.111.2.368

Battista, M. T. (2011). Conceptualizations and Issues Related to Learning Progressions, Learning Trajectories, and Levels of Sophistication. *The Maths. Enthusiast* 8 (3), 507–570. doi:10.54870/1551-3440.1228

Byrge, L., Smith, L. B., and Mix, K. S. (2014). Beginnings of Place Value: How Preschoolers Write Three-Digit Numbers. *Child. Dev.* 85 (2), 437–443. doi:10.1111/cdev.12162

Camos, V. (2008). Low Working Memory Capacity Impedes Both Efficiency and Learning of Number Transcoding in Children. *J. Exp. Child. Psychol.* 99 (1), 37–57. doi:10.1016/j.jecp.2007.06.006

Chan, W. W. L., Au, T. K., and Tang, J. (2014). Strategic Counting: A Novel Assessment of Place-Value Understanding. *Learn. Instruction* 29, 78–94. doi:10.1016/j.learninstruc.2013.09.001

Clements, D. H., and Sarama, J. (2004). Learning Trajectories in Mathematics Education. *Math. Thinking Learn.* 6 (2), 81–89. doi:10.1207/s15327833mtl0602_1

Clements, D. H., and Sarama, J. (2014). "Learning Trajectories: Foundations for Effective, Research-Based Education," in *Learning over Time: Learning Trajectories in Mathematics Education* (Charlotte, NC, US: IAP Information Age Publishing), 1–30.

Cobb, P., and Wheatley, G. (1988). Children's Initial Understandings of Ten. *Focus Learn. Probl. Maths.* 10 (3), 1–26. doi:10.2307/30035059

Dehaene, S. (1992). Varieties of Numerical Abilities. *Cognition* 44 (1–2), 1–42. doi:10.1016/0010-0277(92)90049-n

Dehaene, S., and Cohen, L. (1995). Towards an Anatomical and Functional Model of Number Processing. *Math. Cogn.* 1 (1), 83–120.

Deloche, G., and Seron, X. (1982). From One to 1: An Analysis of a Transcoding Process by Means of Neuropsychological Data. *Cognition* 12 (2), 119–149. doi:10.1016/0010-0277(82)90009-9

Desoete, A., and Grégoire, J. (2006). Numerical Competence in Young Children and in Children with Mathematics Learning Disabilities. *Learn. Individual Differences* 16 (4), 351–367. doi:10.1016/j.lindif.2006.12.006

Dietrich, J. F., Huber, S., Dackermann, T., Moeller, K., and Fischer, U. (2016). Place-value Understanding in Number Line Estimation Predicts Future Arithmetic Performance. *Br. J. Dev. Psychol.* 34 (4), 502–517. doi:10.1111/bjdp.12146

Dowker, A., and Nuerk, H. C. (2016). Editorial: Linguistic Influences on Mathematics. *Front. Psychol.* 7, 1035. doi:10.3389/fpsyg.2016.01035

Fritz, A., Ehlert, A., Ricken, G., and Balzer, L. (2017). *Mathematik und Rechenkonzepte im ersten Schuljahr–Diagnose (MARKO-D1)*. Göttingen: Hogrefe.

Fuson, K. C., Smith, S. T., and Lo Cicero, A. M. (1997a). Supporting Latino First Graders' Ten-Structured Thinking in Urban Classrooms. *J. Res. Maths. Edu.* 28 (6), 738–766. doi:10.5951/jresematheduc.28.6.0738

Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., et al. (1997b). Children's Conceptual Structures for Multidigit Numbers and Methods of Multidigit Addition and Subtraction. *J. Res. Maths. Edu.* 28 (2), 130–162. doi:10.5951/jresematheduc.28.2.0130

Geary, D. C., Hoard, M. K., and Hamson, C. O. (1999). Numerical and Arithmetical Cognition: Patterns of Functions and Deficits in Children at Risk for a Mathematical Disability. *J. Exp. Child. Psychol.* 74 (3), 213–239. doi:10.1006/jecp.1999.2515

Geary, D. C. (2004). Mathematics and Learning Disabilities. *J. Learn. Disabil.* 37, 4–15. doi:10.1177/00222194040370010201

Gilmore, C., Göbel, S. M., and Inglis, M. (2018). *An Introduction to Mathematical Cognition*. London ; New York: Routledge,Taylor and Francis Group.

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H. C. (2014a). Language Affects Symbolic Arithmetic in Children: The Case of Number Word Inversion. *J. Exp. Child. Psychol.* 119, 17–25. doi:10.1016/j.jecp.2013.10.001

Göbel, S. M., Watson, S. E., Lervåg, A., and Hulme, C. (2014b). Children's Arithmetic Development: it Is Number Knowledge, Not the Approximate Number Sense, that Counts. *Psychol. Sci.* 25 (3), 789–798. doi:10.1177/0956797613516471

Herzog, M., Ehlert, A., and Fritz, A. (2017). A Competency Model of Place Value Understanding in South African Primary School Pupils. *Afr. J. Res. Math. Sci. Tech. Edu.* 21 (1), 37–48. doi:10.1080/18117295.2017.1279453

Herzog, M., Ehlert, A., and Fritz, A. (2019). "Development of a Sustainable Place Value Understanding," in *International Handbook of Mathematical Learning Difficulties*. Editors A. Fritz, V. G. Haase, and P. Räsänen (Cham: Springer International Publishing), 561–579. doi:10.1007/978-3-319-97148-3_33

Herzog, M., and Fritz, A. (2020). "Konzeptuelles Stellenwertverständnis als Prädiktor für Rechenfertigkeiten," in *Beiträge Zum Mathematikunterricht 2020*. Editors H.-S. Siller, W. Weigel, and J. F. Wörler (Münster: WTM-Verlag), 421–424.

Herzog, M., and Fritz, A. (20192019). "Validation of a Developmental Model of Place Value Concepts,". *Proceedings of the 43rd Conference of the International*

Group for the Psychology of Mathematics Education. Editors M. Graven, H. Venkat, A. Essien, and P. Vale (Pretoria, South Africa: PME), Vol. 2, 352–359.

Hiebert, J., and Lefevre, P. (1986). “Procedural and Conceptual Knowledge,” in *Conceptual and Procedural Knowledge: The Case of Mathematics* (New York, 1–27.

Houdement, C., and Tempier, F. (2019). Understanding Place Value with Numeration Units. *ZDM Maths. Edu.* 51 (1), 25–37. doi:10.1007/s11858-018-0985-6

Imbo, I., Vanden Bulcke, C., De Brauwer, J., and Fias, W. (2014). Sixty-four or Four-And-Sixty? the Influence of Language and Working Memory on Children’s Number Transcoding. *Front. Psychol.* 5, 313. doi:10.3389/fpsyg.2014.00313

Imbo, I., Vandierendonck, A., and Vergauwe, E. (2007). The Role of Working Memory in Carrying and Borrowing. *Psychol. Res.* 71 (4), 467–483. doi:10.1007/s00426-006-0044-8

Klein, E., Bahnmueller, J., Mann, A., Pixner, S., Kaufmann, L., Nuerk, H. C., et al. (2013). Language Influences on Numerical Development-Inversion Effects on Multi-Digit Number Processing. *Front. Psychol.* 4, 480. doi:10.3389/fpsyg.2013.00480

Klein, E., Moeller, K., Dressel, K., Domahs, F., Wood, G., Willmes, K., et al. (2010). To Carry or Not to Carry-Iis This the Question? Disentangling the Carry Effect in Multi-Digit Addition. *Acta Psychol. (Amst)* 135 (1), 67–76. doi:10.1016/j.actpsy.2010.06.002

Lambert, K., and Moeller, K. (2019). Place-value Computation in Children with Mathematics Difficulties. *J. Exp. Child. Psychol.* 178, 214–225. doi:10.1016/j.jecp.2018.09.008

Lee, M. D., and Sarnecka, B. W. (2009). A Model of Knower-Level Behavior in Number-Concept Development. *Cogn. Sci.* 34 (1), 51–67. doi:10.1111/j.1551-6709.2009.01063.x

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

Mark, W., and Dowker, A. (2015). Linguistic Influence on Mathematical Development Is Specific rather Than Pervasive: Revisiting the Chinese Number Advantage in Chinese and English Children. *Front. Psychol.* 6, 203. doi:10.3389/fpsyg.2015.00203

Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool Origins of Cross-National Differences in Mathematical Competence: The Role of Number-Naming Systems. *Psychol. Sci.* 6 (1), 56–60. doi:10.1111/j.1467-9280.1995.tb00305.x

Moeller, K., Klein, E., and Nuerk, H. C. (2011b). Three Processes Underlying the Carry Effect in Addition-Eevidence from Eye Tracking. *Br. J Psychol* 102 (3), 623–645. doi:10.1111/j.2044-8295.2011.02034.x

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H. C. (2011a). Early Place-Value Understanding as a Precursor for Later Arithmetic Performance-Aa Longitudinal Study on Numerical Development. *Res. Dev. Disabil.* 32 (5), 1837–1851. doi:10.1016/j.ridd.2011.03.012

Moura, R., Lopes-Silva, J. B., Vieira, L. R., Paiva, G. M., Prado, A. C., Wood, G., et al. (2015). From “Five” to 5 for 5 minutes: Arabic Number Transcoding as a Short, Specific, and Sensitive Screening Tool for Mathematics Learning Difficulties. *Arch. Clin. Neuropsychol.* 30 (1), 88–98. doi:10.1093/arclin/acu071

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2013). Transcoding Abilities in Typical and Atypical Mathematics Achievers: The Role of Working Memory and Procedural and Lexical Competencies. *J. Exp. Child. Psychol.* 116 (3), 707–727. doi:10.1016/j.jecp.2013.07.008

Nuerk, H. C., Weger, U., and Willmes, K. (2001). Decade Breaks in the Mental Number Line? Putting the Tens and Units Back in Different Bins. *Cognition* 82 (1), B25–B33. doi:10.1016/S0010-0277(01)00142-1

Nuerk, H.-C., Moeller, K., and Willmes, K. (2014). *Multi-digit Number Processing*. Editors R. Cohen Kadosh and A. Dowker (Oxford University Press), Vol. 1. doi:10.1093/oxfordhb/9780199642342.013.021

Nührenbörger, M., and Steinbring, H. (2008). “Manipulatives as Tools in Teacher Education,”. *International Handbook of Mathematics Teacher Education.* Editors S. Llinares and O. Chapman (Leiden, Netherlands: Brill | Sense), 2, 157–181. doi:10.1163/9789087905460_009

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H. C., and Moeller, K. (2011). One Language, Two Number-word Systems and many Problems: Numerical Cognition in the Czech Language. *Res. Dev. Disabil.* 32 (6), 2683–2689. doi:10.1016/j.ridd.2011.06.004

Pixner, S., Nuerk, H.-C., and Moeller, K. (2016). Einflüsse der Inversion auf die Verarbeitung mehrstelliger Zahlen bei deutschsprachigen Kindern - ein Überblick. *Lernen und Lernstörungen* 5 (3), 173–188. doi:10.1024/2235-0977/a000021

Poncin, A., Van Rinsveld, A., and Schiltz, C. (2020). Units-first or Tens-First: Does Language Matter when Processing Visually Presented Two-Digit Numbers? *Q. J. Exp. Psychol. (Hove)* 73 (5), 726–738. doi:10.1177/1747021819892165

Reiss, K., and Obersteiner, A. (2019). “Competence Models as a Basis for Defining, Understanding, and Diagnosing Students’ Mathematical Competences,” in *International Handbook of Mathematical Learning Difficulties*. Editors A. Fritz, V. G. Haase, and P. Räsänen (Cham: Springer International Publishing), 43–56. doi:10.1007/978-3-319-97148-3_4

Ricken, G., Fritz, A., and Balzer, L. (2013). *MARKO-D: Mathematik-und Rechenkonzepte im Vorschulalter-Daignose: Mit Ergänzungstest Konventions-und Regelwissen*. Hogrefe.

Rittle-Johnson, B., and Schneider, M. (2015). *Developing Conceptual and Procedural Knowledge of Mathematics*. Oxford Handbook of Numerical Cognition, 1118–1134.

Ross, S. H. (1989). Parts, Wholes, and Place Value: A Developmental View. *Arithmetic Teach.* 36 (6), 47–51. doi:10.5951/at.36.6.0047

Scherer, P., and Moser Opitz, E. (2010). *Fördern im Mathematikunterricht der Primarstufe*. Heidelberg: Spektrum Akademischer Verlag.

Schulz, A. (2014). *Fachdidaktisches Wissen von Grundschullehrkräften*. Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-08693-0

Selter, C. (2001). Addition and Subtraction of Three-Digit Numbers: German Elementary Children’s Success, Methods and Strategies. *Educ. Stud. Maths.* 47 (2), 145–173. doi:10.1023/a:1014521221809

Siegler, R. S., and Alibali, M. W. (2005). *Children’s Thinking*. 4th ed. Upper Saddle River, N.J: Pearson Education/Prentice Hall.

Steffe, L. P. (1992). Schemes of Action and Operation Involving Composite Units. *Learn. Individual Differences* 4 (3), 259–309. doi:10.1016/1041-6080(92)90005-Y

Van de Walle, J., Karp, K., and Bay-Williams, J. (2016). in *Elementary and Middle School Mathematics: Teaching Developmentally*. Editor M. Fossel. Ninth (USA: Pearson Education).

van der Ven, S. H. G., Klaiber, J. D., and van der Maas, H. L. J. (2017). *Four And Twenty* Blackbirds: How Transcoding Ability Mediates the Relationship between Visuospatial Working Memory and Math in a Language with Inversion. *Educ. Psychol.* 37 (4), 487–505. doi:10.1080/01443410.2016.1150421

Zuber, J., Pixner, S., Moeller, K., and Nuerk, H. C. (2009). On the Language Specificity of Basic Number Processing: Transcoding in a Language with Inversion and its Relation to Working Memory Capacity. *J. Exp. Child. Psychol.* 102 (1), 60–77. doi:10.1016/j.jecp.2008.04.003

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

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