This is a repository copy of *Domain-Specific and Domain-General Skills as Predictors of Arithmetic Fluency Development*.

Article:
Mononen, Riikka, Korhonen, Johan, Hgeland, Karoline et al. (3 more authors) (2025) Domain-Specific and Domain-General Skills as Predictors of Arithmetic Fluency Development. Learning and individual differences. 102585. ISSN 1041-6080

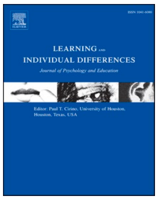
https://doi.org/10.1016/j.lindif.2024.102585

Learning and Individual Differences 117 (2025) 102585

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif



Domain-specific and domain-general skills as predictors of arithmetic fluency development

Riikka Mononen [a,b,*], Johan Korhonen [c], Karoline Hægeland [b], Matin Younesi [b], Silke M. Göbel [d,b], Markku Niemivirta [e]

[a] *Special and Inclusive Education, University of Oulu, Finland*
[b] *Department of Special Needs Education, University of Oslo, Norway*
[c] *Faculty of Education and Welfare Studies, Åbo Akademi University, Finland*
[d] *Department of Psychology, University of York, UK*
[e] *School of Applied Educational Science and Teacher Education, University of Eastern Finland, Finland*

ARTICLE INFO

Keywords:
Arithmetic fluency
Number sequence skills
Gender
Symbolic magnitude processing
Working memory

ABSTRACT

We investigated Norwegian children's ($n = 262$) development in arithmetic fluency from first to third grade. Children's arithmetic fluency was measured at four time points, domain-specific (i.e., symbolic magnitude processing and number sequences) and domain-general skills (i.e., working memory, rapid naming, non-verbal reasoning, and sustained attention) once in the first grade. Based on a series of growth mixture models, one developmental trajectory best described the data. Multigroup latent growth curve models showed that girls and boys developed similarly in their arithmetic fluency over time. Symbolic magnitude processing and number sequence skills predicted both initial level and growth in arithmetic fluency, and working memory predicted only initial level, similarly for boys and girls. Mother's education level predicted the initial level of arithmetic fluency for boys, and rapid naming predicted growth for girls. Our findings highlight the role of domain-specific skills in the development of arithmetic fluency.

1. Introduction

Fluent arithmetic fact retrieval is one of the key areas in mathematics that children need to learn during their first years in school, as this skill is the foundation for later, more advanced mathematics learning (Xu et al., 2021). Knowing the answer by heart for addition and subtraction facts, such as for $2 + 7$ or $13–9$, will free up cognitive processing resources for other, more demanding parts of the math problem to be solved (e.g., in multi-digit calculations or in word problem solving). There are large individual differences in children's mathematical skills at the beginning of schooling (Aunio & Niemivirta, 2010), and arithmetic skills are not an exception (Vanbinst et al., 2015; Xu et al., 2021). From a research methodological viewpoint, person-centered approaches seem particularly well suited to capture individual differences in development (Hickendorff et al., 2018), and these have started to emerge also in the research field of mathematical cognition (Bakker et al., 2023; Scalise et al., 2021). The advantage of this data-driven approach is that (1) not all students are assumed to follow a single developmental trajectory, and (2) artificial cut-off points, such as performing below the 25th percentile, are not needed to group students. In this study, a person-centered approach is used to explore what kind of developmental trajectories of arithmetic fluency can be identified in children from first to third grade.

In many studies children's mathematics performance and development have been associated with certain domain-specific (e.g., non-symbolic and symbolic magnitude processing, mathematical vocabulary) and domain-general (e.g., working memory, language, fluid intelligence) factors (for an overview, see De Smedt, 2022). Fewer studies, especially longitudinal ones, have included simultaneously several domain-specific and domain-general factors as predictors (for some exceptions, see Bakker et al., 2023; Banfi et al., 2024; Fuchs et al., 2010; Geary, 2011a; Träff et al., 2020, 2023; Zhang et al., 2017). Inclusion of different factors simultaneously would give a more comprehensive and accurate picture of each predictor's unique contribution to mathematics development. In other words, using only a few factors as predictors, may easily overestimate their effects when other important factors for development are not considered. Further, rather few studies have simultaneously investigated how both domain-specific and domain-

---

* Corresponding author at: P.O. Box 8000, 90014 University of Oulu, Finland.
E-mail address: riikka.mononen@oulu.fi (R. Mononen).

https://doi.org/10.1016/j.lindif.2024.102585
Received 23 December 2023; Received in revised form 12 September 2024; Accepted 9 November 2024

general factors predict the profile membership in mathematics achievement or development (Bakker et al., 2023; Scalise et al., 2021), and to our knowledge, none concerning arithmetic fluency development.

Here, we expand prior research by investigating simultaneously several domain-specific (i.e., symbolic magnitude processing and number sequence skills) and domain-general factors (i.e., non-verbal reasoning, working memory, rapid automatized naming [RAN], sustained attention), and together with demographics (i.e., age and mother's educational level), with the aim to find out how these predict arithmetic fluency development. Although recent research indicates that there are minimal gender differences in early mathematics performance (Hyde et al., 2008; Lachance & Mazzocco, 2006), less is known whether gender moderates arithmetic fluency development (Martens et al., 2011), as well as the effects of these predictors on arithmetic fluency development.

1.1. The development of arithmetic fluency

Acquiring different calculation strategies characterize the development of foundational addition and subtraction skills (Geary et al., 2004; Jordan et al., 2003). Various counting procedures are initially used to solve basic arithmetic problems, such as $3 + 5$ (Koponen et al., 2007). In the beginning, the child counts both addends to find the answer, before they steadily shift towards more advanced counting procedures, such as counting on from the larger addend (Fuson, 1992). The counting-on procedure requires that children identify the larger addend, and thus draws on their understanding of numerical magnitude (De Smedt, Verschaffel, & Ghesquière, 2009). Gradually, by using these counting strategies multiple times, children start to build associations between the problem and the answer, and arithmetic facts (e.g., $3 + 5 = 8$) are finally stored in long-term memory (Siegler & Shrager, 1984). The calculation strategies, which the child uses, change over time, but both procedural strategies and fact retrieval continue to exist throughout childhood and into adulthood (Geary et al., 2004; Jordan et al., 2003).

Most of the children develop in their learning of addition and subtraction facts so that by the age of nine, they can retrieve the addition and subtraction facts fluently (i.e., quickly and accurately) in the number range between 1 and 20 (Aunio & Räsänen, 2016). However, some children struggle to retrieve the facts even after years of practice (Jordan et al., 2003; Träff et al., 2020). They may find the correct answer for the problem, but need more time, if still using procedural strategies as their main calculation strategy. Difficulties in arithmetic fact retrieval have been found to characterize children with mathematical learning difficulties (Geary, 2011b; Vanbinst et al., 2014). Regardless of their performance level in arithmetic skills, children's development seems to be rather stable across the early grades (Sorvo et al., 2019; Xu et al., 2021). In other words, those who start with a lower set of arithmetic skills tend to remain performing lower over the years compared to their typically performing peers, and this achievement gap may even widen over the years (Aunola et al., 2004).

1.2. The role of domain-specific skills in arithmetic fluency development

Domain-specific factors in the context of mathematics learning typically refer to mathematical skills (e.g., symbolic magnitude processing), but also other mathematics related factors, such as mathematical home environment or mathematical vocabulary (De Smedt, 2022). In this study, we focus on two mathematical skills relevant for this age group, symbolic magnitude processing and number sequence skills. Both skills involve using numbers at symbolic level either as numerals (e.g., 5) or number words (e.g., five), respectively. Symbolic magnitude processing skills tap more the cardinal aspect of number (i.e., a child needs to understand the magnitudes of numbers when comparing two numbers) while number sequence skills the ordinal aspect of number (i.e., a child needs to understand the relative position of numbers in a sequence when reciting number words from various points in a number sequence) (Devlin et al., 2022). Both skills have also been found in prior studies to be associated with arithmetic fluency (Vanbinst et al., 2016; Zhang et al., 2017, 2020) or more generally with mathematics achievement (Brankaer et al., 2017; Koponen et al., 2019; Xenidou-Dervou et al., 2018), but rarely been investigated simultaneously as predictors for arithmetic development (however, see Banfi et al., 2024). Recently, Finke et al. (2022) suggested that there may be a shift during the first school years from symbolic magnitude processing to order processing becoming a better predictor of arithmetic performance, which makes inclusion of both symbolic magnitude processing and number sequence skills as predictors of arithmetic fluency interesting.

Symbolic magnitude processing is considered as one of the key foundational number processing skills, often considered to belong under a wider concept of number sense (De Smedt et al., 2013). Symbolic magnitude processing is typically assessed with tasks that involve Arabic digits, and children are asked to respond as quickly as possible, which one of two numbers is the larger one (Brankaer et al., 2017). Good symbolic magnitude processing skills are characterized by fast and accurate responses, and it has been found to be a strong predictor of children's mathematics achievement across different grades in elementary school (Brankaer et al., 2017; Holloway & Ansari, 2009), and of approximate arithmetic performance (Wei, Deng, et al., 2018), as well as arithmetic fluency (Banfi et al., 2024; Träff, 2013; Vanbinst et al., 2016). In their study, Banfi et al. (2024) found magnitude processing skills, assessed with dots and digits comparison tasks in grade 1, to predict arithmetic fluency in grades 2 and 3. In contrast, Träff et al. (2020) did not find symbolic magnitude processing skills measured in early grades (K–3) to predict math performance in grade 6, although the sixth graders with low mathematics performance showed weaker symbolic magnitude processing skills compared to their peers during all years from kindergarten to grade 3. The shared mechanisms that link symbolic magnitude processing and arithmetic fluency together, are still partly unclear. One explanation might be that in both tasks, as how those are measured, the child needs first to quickly access the magnitudes and/or verbal counterparts of two numbers (Habermann et al., 2020), before processing them further, that is, comparing them or solving the answer for the arithmetic fact. Vanbinst et al. (2014) found that students with persistent mathematical learning difficulties, specifically, with a fact retrieval deficit, also showed weak symbolic magnitude processing skills. In their later study (Vanbinst et al., 2016), its role was discovered to be as important for arithmetic as phonological awareness is for reading.

Children's counting skills start to emerge in their early childhood through play and practice, and children gradually learn the numerical symbols of their culture. This includes number words (e.g., one, two, three) and numerals (e.g., 1, 2, 3), which can be compared to one another and ordered (Scalise & Ramani, 2021). It takes a surprisingly long time for children to understand the meaning of number words, and even though they can count from 1 to 10 accurately, they might spend a year or more to understand the relative magnitudes of the numbers they count (Le Corre & Carey, 2007). Reciting number sequences forwards and backwards has shown to be a strong predictor of arithmetic performance (Gilmore & Batchelor, 2021; Träff et al., 2023) and development (Koponen et al., 2019; Zhang et al., 2017). This is a rather expected finding, as early calculation strategies rely much on counting procedures (e.g., count all and count on strategies) (Fuson, 1992). However, the findings may partly depend on the type of the arithmetic measure used (e.g., 1- or 2-digit problems; verbal or written tasks), as well as what other cognitive covariates are included in the study. For example, Träff et al. (2023) found that counting skills were related to the initial arithmetic performance, but they did not predict arithmetic development, while controlling for non-symbolic magnitude processing, spatial processing, non-verbal and verbal reasoning, and working memory. Furthermore, Banfi et al. (2024) found forward number sequence skills (i.e., count up from 1) measured in grade 1 to predict only third grade arithmetic fluency in quantile 16 (the lowest performing group), but not arithmetic fluency in grades 1 and 2.

1.3. *The role of domain-general skills in arithmetic development*

Domain-general skills are those that are important for learning in general. In this study, we chose to include a set of domain-general skills that, in prior studies, were found to be associated specifically with arithmetic performance: non-verbal reasoning, working memory (i.e., central executive), rapid automatized naming (RAN), and sustained attention.

Non-verbal reasoning skills, also referred to as fluid intelligence (Geary et al., 2012), have been reported to predict mathematics performance (Fuchs et al., 2010; Geary, 2011a; Kyttälä & Lehto, 2008; Pina et al., 2014; Träff et al., 2020; Xenidou-Dervou et al., 2018), and arithmetic performance and development (Banfi et al., 2024; Cai et al., 2018; Geary et al., 2012; Träff et al., 2023; Wei, Guo, et al., 2018) in different age groups. Banfi et al. (2024) found nonverbal reasoning measured at grade 1 to predict arithmetic fluency in grade 2 (at quantile levels 16 and 50, the lowest performing and average performing group, respectively) and in grade 3 (at quantile 16). The type of the arithmetic task, and the process of solving it, seems to play a role in this relation. In their study, Wei, Guo, et al. (2018) found that non-verbal IQ predicted intercept and slope of arithmetic accuracy from second to fifth grade, but not arithmetic fluency. Geary et al. (2012) did not find non-verbal reasoning affecting arithmetic fact retrieval directly, but to contribute more to the development of using other calculation strategies, such as decomposition, from first to fourth grade. This is a quite expected finding. When directly retrieving a fact from long-term memory, the child does not need to pay much attention to reasoning and making logical decisions while proceeding in the task, as compared to, for example, in decomposing an arithmetic problem (e.g., $7 + 6 = 7 + 3 + 3$) or in mathematical problem solving (Engle, 2018). Träff et al. (2023), in contrast, found that non-verbal reasoning was linked to arithmetic performance of six-year-olds, but not to the development of arithmetic performance over one school year. However, in their study, arithmetic performance measures included also non-timed, verbally presented arithmetic problems (e.g., $17 + 44$ or 85–28). In sum, concerning arithmetic fact fluency, it seems likely that non-verbal reasoning has relatively little effect on this skill, but it contributes more when a child uses other calculation strategies than fact retrieval.

Working memory has extensively been studied and found to be associated with mathematics and arithmetic performance (for meta-analyses, see Chen & Bailey, 2021; Friso-van den Bos et al., 2013). The component of central executive (CE), involved in controlling and processing of information, has been found to be concurrently linked to (Allen et al., 2020; Banfi et al., 2024; Xenidou-Dervou et al., 2018; however, for contrasting findings, see Fuchs et al., 2006) or to predict mathematics achievement (De Smedt, Janssen, et al., 2009; Fuchs et al., 2010; Geary, 2011a), and specifically arithmetic fluency (Andersson, 2008; Barrouillet & Lépine, 2005; Geary et al., 2012; Hoff et al., 2023; Wei, Guo, et al., 2018). A higher CE capacity eases the learning of the associations between the calculation problem and the answer and enables the activation of these associations and inhibition of irrelevant information, when retrieving the facts from the long-term memory (Barrouillet & Lépine, 2005). Studies by Barrouillet and Lépine (2005) and Geary et al. (2012) showed that children in early grades with a higher CE capacity were using fact retrieval in addition problems more frequently compared to their peers with a lower CE capacity. However, in their longitudinal study Geary et al. (2012) found that this finding was limited to the first grade only, and diminished from first to fourth grade once intelligence and in-class attentive behavior were controlled for. Further support for the relation between the CE and arithmetic skills comes from Wei, Guo, et al. (2018) and Andersson (2008). Following students from second to fifth grade, Wei, Deng, et al. (2018) and Wei, Guo, et al. (2018) found that CE (measured as inhibition and shifting) predicted arithmetic fluency intercept, but not growth, while controlling for nonverbal IQ, speed of processing, working memory, and number sense. Andersson (2008) found that CE was uniquely linked to arithmetic fluency of third and fourth graders, while also considering IQ, reading skills, and other working memory components. Furthermore, Hoff et al. (2023) found that CE measured at pre-school, but not at the first grade, predicted third grade arithmetic fluency, while accounting also for RAN, number knowledge, and processing speed. These findings suggest that CE is a significant contributor to arithmetic fact fluency, while its unique role at different ages seems to still require further investigation, due to somewhat contradicting and a rather limited number of longitudinal studies at present.

Rapid automatized naming (RAN), that is quickly naming of familiar colors or objects (i.e., non-alphanumeric RAN) or letters and numbers (i.e., alphanumeric RAN), has been linked to and found to predict arithmetic fluency in early grades (Hornung et al., 2017; for a meta-analysis, see Koponen et al., 2017; Wang, 2020), while also controlling for number knowledge, processing speed, and working memory (Hoff et al., 2023). The reason for this relationship most likely stems from the need of accessing and retrieving phonological presentations (e.g., “green” or “seven”) for a stimulus from the long-term memory (Koponen et al., 2017). Further, Pulkkinen et al. (2022) showed that those third graders, who were dysfluent in arithmetic, had lower RAN performance during the early grades, compared to their peers who were fluent in arithmetic.

Sustained attention is one of the primary elements of attention, and it enables the child to maintain vigilance, select, and focus attention over time (Cohen, 2011). Its relation to mathematics and arithmetic performance and development is still rather sparsely investigated (Orbach & Fritz, 2022). The few studies have found sustained attention to be related to arithmetic skills both when sustained attention was measured with performance-based tests (Orbach & Fritz, 2022) and as experimenter ratings (West et al., 2021). However, Szűcs et al. (2014) found sustained attention to be related to both non-symbolic and symbolic magnitude processing but not to mathematics achievement.

1.4. *Boys and girls are more similar than different in cognitive functions*

Meta-analytical research has shown that boys and girls are more similar than different in most cognitive functions (Hyde, 2016), and if any differences have been found in individual studies, their effects are typically rather small. In early grades, rather minimal or nonexistent gender differences have been found in overall mathematics performance (Hyde et al., 2008; Lachance & Mazzocco, 2006). Regarding arithmetic performance, Xu et al. (2021) found that boys slightly outperformed girls in subtraction facts in the second grade, while Lachance and Mazzocco (2006), in contrast, found girls to solve more accurately timed addition and subtraction facts. When including a sample of students from grades 1–9, Martens et al. (2011) found that gender differences in arithmetic performance, especially in addition and subtraction, started to emerge from Grade 6 onward in favor of boys. These differences were evidenced to be larger in higher than low-performing groups. Similarly, Räsänen et al. (2021) found a small advantage for boys in arithmetic fluency in a large sample of Finnish students from grades 3–9. Furthermore, boys were overrepresented in both ends of the ability continuum.

Most previous studies have focused on mean level differences in arithmetic performance. Studies on gender differences focusing on the relationship between domain-general, domain-specific factors and arithmetic performance are scarce with the exception of motivational research that has found some gendered pathways between math interest, self-concept, educational aspirations and performance (e.g., Cvencek et al., 2021; Widlund et al., 2024). In one of the few studies to date, Carr et al. (2008) found that the type of strategy used differentially predicted arithmetic performance for girls and boys. Higher performance was related to cognitive strategy use for boys while higher performance was related to manipulative strategy use for girls. Rosselli et al. (2009) found similar cognitive correlates with math performance for boys and girls, but they found that spatial ability could explain small gender differences in favor of boys in arithmetic performance.

Regarding gender differences in non-verbal reasoning, the results are mixed, partly depending on the measure utilized. A meta-analysis (Lynn & Irwing, 2004) for studies using Raven's Progressive Matrices as a measure for non-verbal reasoning found no gender differences until after the age of 15, however, a small effect ($d = 0.21$) in favor of boys was found in children aged 5–11 years old, when Raven's Colored Progressive Matrices was used. Further, Xu et al. (2021) found that girls outperformed boys in non-verbal reasoning in the second grade ($d = 0.34$), while Lachance and Mazzocco (2006) and Toivainen et al. (2017) found no gender differences in early grades.

Regarding gender differences in children's working memory, we must rely on some independent research work, as there is a lack of a recent review or a meta-analysis. The findings are contradicting. While Gray et al. (2015) found no gender differences in working memory among 5–9-year-olds, Pelegrina et al. (2015) and Vuontela et al. (2003), using n-back tasks as a measure of working memory, identified girls to outperform boys in accuracy, while boys being faster responders. This finding is suggested to reflect a larger degree of immaturity in boys than girls in childhood (Vuontela et al., 2003).

Gender differences in RAN seem to partly differ depending on the type of RAN test used. Among primary school students, Lachance and Mazzocco (2006) found girls to outperform boys only on Colors subtest, but not on Object, Numbers, or Letters. Current studies concerning RAN and mathematics performance miss reporting possible gender differences (Hoff et al., 2023; Koponen et al., 2017, 2016; Wang, 2020).

Gender differences in sustained attention have been less studied in this age group, although attentional issues have been in focus among the children with a diagnosis of ADHD (for a meta-analysis, see Hasson & Fine, 2012). The findings from a few studies are inconsistent but may partly be explained using different types of measures (e.g., Continuous Performance Test vs. Psychomotor Vigilance Task). Sussman and Tasso (2013) found that girls outperformed boys, both in making less errors and being faster in completing the tasks. However, here, the within-gender differences in performance far exceeded the between-gender variation. In contrast, Efrat and Orna (2022) and Venker et al. (2007) found boys being faster in their responses, although gender differences decreased with age, and disappeared around the age of 11 (Venker et al., 2007). In our study, the measure of sustained attention is more similar to the measure of Continuous Performance Test, which Sussman and Tasso (2013) used in their study.

1.5. *Current study*

As previous studies have mainly investigated the development of children's arithmetic fluency from a variable-centered perspective (e.g., Aunola et al., 2004; Xu et al., 2021), in this study we approach arithmetic development methodologically from the person-centered approach. This means that we aim to find out if there is more than one developmental trajectory, which may reveal different patterns and growths in children's arithmetic development, giving us a more comprehensive view of development. Based on previous research (e.g., Zhang et al., 2020), we can expect to find at least two to three different trajectories. Furthermore, prior studies have examined and found how certain domain-specific and domain-general factors are related to or predict the development of arithmetic fluency, but the studies have varied widely in the factors they have included simultaneously, and the results have also been somewhat mixed. Including multiple domain-specific and domain-general factors allows us to see which of these factors remain the strongest predictors while controlling for others, for example, do domain-general factors lose their predictive power when considered alongside domain-specific factors. In terms of domain-specific factors, both symbolic magnitude processing (Banfi et al., 2024; Träff, 2013; Vanbinst et al., 2016) and number sequence skills (Koponen et al., 2019; Zhang et al., 2017) have been shown to be strong predictors of the development of arithmetic fluency in the early grades. However, they have rarely been treated simultaneously as predictors in the same study. Although all of the domain-general factors chosen for this study, namely nonverbal reasoning, working memory (CE), RAN, and sustained attention, have been shown to be related to or predictive of arithmetic performance to varying degrees, CE and RAN in particular appear to have a strong association with arithmetic performance and development, even when controlling for domain-specific mathematical skills (e.g., Hoff et al., 2023). Moreover, the influence of predictors seem to vary depending on the type of arithmetic outcome measures used (e.g., accuracy or fluency measures) and whether the focus was on performance level versus growth (e.g., Träff, 2013; Wei, Guo, et al., 2018).

In this longitudinal study, we trace children's arithmetic fluency development from first to third grade. We extend the current state of research by exploring if we can identify distinct developmental trajectories using a person-centered approach. Further, we aim to find out which domain-specific (i.e., symbolic magnitude processing and number sequence skills), domain-general (i.e., non-verbal reasoning, working memory, RAN, sustained attention), and demographic factors (i.e., age and mother's educational level), when investigated simultaneously, best predict the development. Further, we look at whether gender moderates (1) the development of arithmetic fluency and (2) the relations between the predictors and arithmetic fluency. We set the following research questions (RQ) and hypotheses (H):

RQ1. What kind of developmental trajectories of arithmetic fluency can be identified in children, and is this development moderated by gender?

Prior research has shown individual differences in children's arithmetic skills (Vanbinst et al., 2015; Xu et al., 2021; Zhang et al., 2020). Therefore, we expect (H1.1) to find more than one developmental trajectory, such as groups of children showing different levels of performance (e.g., low, average, and high) in the first grade. We expect children to improve their skills from first to third grade, and those who start with weaker skills, will remain performing weaker, and vice versa (Sorvo et al., 2019; Xu et al., 2021). Further, we expect to confirm (H1.2) that gender has no effect on arithmetic development at this age group, referring to rather non-existing gender differences in prior findings in overall early mathematics performance (Hyde et al., 2008; Lachance & Mazzocco, 2006).

RQ2. How do domain-specific, domain-general, and demographic factors (age and mother's educational level) predict the development of arithmetic fluency, and does gender moderate these predictions?

All the domain-specific and domain-general skills included in our study have been shown in prior studies to influence arithmetic fluency. We expect (H2.1) to find effects of all domain-specific and domain-general skills on the intercept of arithmetic fluency. While longitudinal findings are still rather scarce and these factors are now simultaneously investigated, we cannot explicitly hypothesize their expected effects on arithmetic fluency development (i.e., slope). We expect these predictors to be similar between the genders (Hyde, 2016) (H2.2), although some individual studies have observed gender differences either in favor of boys or girls, for example girls performing better than boys in RAN of colors (Lachance & Mazzocco, 2006) and in sustained attention (Sussman & Tasso, 2013). Of demographic predictors, we do not expect age and mother's educational level to have a strong effect, given the fact that our sample is from one grade, and further, from a Nordic country, where all children have equal opportunities for schooling in public schools (Aunio & Niemivirta, 2010).

2. Method

2.1. *Participants and procedure*

The study is part of the larger longitudinal research project iSeeNumbers, and parts of the data have been used in prior studies (e.g., Rawlings et al., 2023), however, the current results represent a substantial contribution on its own. The participants were 262 Norwegian children (M_{age} = 6 y. 9 m., SD = 3.33 m., 46.1 % girls in the beginning of the study), from 12 classrooms in five schools in the capital region of Norway. Prior to data collection, ethical approval was applied for and granted by the Norwegian Centre for Research Data, and consent for participation was obtained from the children's parents and teachers.

The Covid-19 pandemic affected the data collection. Initially, there were five time points planned for data collection, biannually in each grade (from first grade spring term 2019 to third grade spring term 2021). Due to lockdown in spring 2020, the initially planned third time point could not be carried out. Also, due to uncertain pandemic times, the ways of collecting data needed to be adjusted. Data collections at t1 (first grade spring) and t2 (second grade autumn) were conducted at the data collection site outside schools as "adventure days". This 4-h day included testing a variety of math and cognitive skills individually and in small groups by trained research assistants. At t3 (third grade autumn) tests were carried out in small groups at schools by research assistants, also taking into account Covid-19 restrictions set by the government and schools. At t4 (third grade spring) Covid-19 set more restrictions for data collection, and research assistants were not allowed to visit schools. The research assistants joined the assessments via Teams, with audio and video connection, while classrooms were supervised by children's teachers. Teachers were given information about the testing situation, and the required testing materials were delivered to schools a few days prior data collection. No technical issues were faced that would have negatively affected the data collection.

There was some missing data, due to Covid-19 restrictions, children absent from days of data collection, or moving away during the study. Regarding arithmetic performance, the attrition was 2.3 % at t1, 7.6 % at t2, 15.6 % at t3, and 20.6 % at t4, and for covariates ranging from 1.9 % (number sequences and attention) to 9.9 % (age and mother's educational level). Little's MCAR test indicated that data was missing completely at random, $\chi 2$ (365) = 408.756, p = .057. Due to the analytical approach in this study (i.e., inclusion of conditional models), missing data was imputed a priori data analyses using the Expectation-Maximization (EM) algorithm in SPSS29 to be able to use all available data in the main analyses.

2.2. Measures

Children's arithmetic skills were measured at all timepoints (t1–t4) using the same measure. Domain-specific and domain-general skills, which were used as predictors, were measured only in the first grade (t1). Demographics of each child were collected with a questionnaire from the parents, in which they reported the gender (coded as girl = 0, boy = 1) and the age of their child, and mother's highest educational level (1 = comprehensive school, 2 = upper secondary school, 3 = Bachelor's degree, 4 = Master's degree, 5 = PhD degree).

2.2.1. Arithmetic fluency

Regnefaktaprøven (Klausen & Reikerås, 2016) is a standardized Norwegian test measuring arithmetic fluency as arithmetic facts. In our study, subtests of addition and subtraction within the number range between 1 and 20 (e.g., 8 + 4 or 11–2), were used at each time point. Each subtest consisted of one page with 45 calculation problems per page, which were either addition or subtraction facts. The child had 2 min to solve as many calculations as possible in each subtest. For each correct answer one point was given and for each incorrect answer zero points. A composite score for arithmetic fluency was created (max. 90 points).

2.2.2. Domain-specific predictors

For domain-specific skills, *symbolic magnitude processing* skills were measured using The SYmbolic Magnitude Processing (SYMP) Test (Brankaer et al., 2017). Here, we use only the one-digit subtest with digits between 1 and 9. The test contains 60-digit pairs that are presented in four columns of 15 pairs. The child was asked to cross out the larger of the two digits (e.g., which is larger, 8 or 9), and was given 30 s to solve as many items as possible. To ensure that the child understands the task, four practice trials were included prior to the test items (see test and item description more in detail: Brankaer et al., 2017). For each correct answer one point was given and for incorrect answers zero points. The number of items that were correctly answered in 30 s was used as a sum score.

To measure children's *number sequence* skills, we used the Number Sequences task from a standardized LukiMat Mathematics test battery (Salminen & Koponen, 2011). The child was asked to recite orally different types of number sequences (counting by 1 s, 2 s, 5 s and 10s) forwards (13 tasks) and backwards (19 tasks) within the number range between 1 and 100. For example, the child was asked to count up by one starting from a given number (e.g., 4, 17, 20 or 74), and to stop once five numbers were produced, or to count up by fives (from number five) until reaching 50. The backward number sequence tasks were similar. One point was given for a correct sequence and zero for an incorrect answer (max. 29 points).

2.2.3. Domain-general predictors

For domain-general skills, *non-verbal reasoning skills* were measured with Raven's Colored Progressive Matrices (Raven et al., 1990). The child needed to find a piece out of six alternatives that fitted into the picture pattern. The tasks were given in a booklet, one on each page, so that the child could proceed in the task at their own pace. There were two practice items in the beginning to make sure that the child understood the task. The maximum score for the whole task was 34 points.

Rapid automatized naming (RAN) was measured using a subtest of naming colors Hurtig Benevning [Rapid naming] from the standardized test battery Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4), for 5–12 years old (Semel et al., 2003). The child was asked to name as fast as possible 36 colored (i.e., yellow, red, green, and blue) circles in a given order. The colored circles were presented on a sheet in six rows of six circles in each. A composite score was calculated: the number of total correct responses was divided by total naming time in seconds and multiplied by ten. The higher this value was, the more items the child could name correctly per second, thus displaying better RAN.

The component of central executive in *working memory* was measured using a subtest of the *Backwards Digit Span* from the standardized Wechsler Intelligence Scale for Children ages 6 through 16 (V Norwegian version) (Wechsler, 2017). In the Backwards Digit Span test the child needed to recall orally 2–8 digits backwards. There was a total of 9 blocks in the test, each including two items. Blocks 1 and 2 included two digits to recall backwards, blocks 3 and 4 three digits, and thereafter the number of digits increased by one so that in block 9 there were eight digits to recall. According to the manual, the test was stopped if both items in the block were incorrect. One point was given for a correct answer and zero for an incorrect answer. The total maximum score was 18 points.

Sustained Attention was measured with the Attention Sustained subtest from the standardized Leiter International Performance Scale – Third Edition (Roid et al., 2013). The child completed four different tasks, each presented on a separate page. In each task the child needed to cross out as many of the pictures as possible that looked like the target picture presented at the top of the page (e.g., the target picture was a square and there were also other figures than squares presented on the page). The child had 30 s time for two of the tasks and 60 s for another two tasks. A composite score of four subtests was calculated by adding all correct responses and then subtracting all incorrect responses.

2.3. Data analysis

Imputation of missing data, descriptive statistics, and correlations of the observed variables were conducted using SPSS29. Regarding identifying the number of trajectory groups, we used MPlus8 version 8.8 (Muthén & Muthén, 1998), first, for the analyses of latent growth curve modeling (LGCM), the one-class model, followed by growth mixture modeling (GMM), iteratively adding classes up to five. For the model fit indices, we used the criteria suggested by Hu and Bentler (1999) and Schermelleh-Engel et al. (2003): the χ^2 statistic p-value $> .05$, comparative fit index (CFI) value >0.95, the root mean square of error approximation (RMSEA) value <0.05, and the standardized root mean squared residual (SRMR) value <0.05. When comparing each model against the previous model, to make the decision on the number of classes, we followed Ferguson et al. (2020), and compared the Bayesian Information Criterion (BIC), Sample Adjusted BIC, and Akaike's Information Criterion (AIC) with a lower value representing a better fit; entropy (i.e., a higher value representing a better fit of profiles to the data), Vuong-Lo-Mendell-Rubin Likelihood Ratio (VLMR) and Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR) (i.e., statistically significant difference between the models indicates more groups to be a more parsimonious model), as well as the number of students in each group (i.e., not including groups with <5 % of the total sample), and finally, having a meaningful relation to the theory.

Note, that the modeling procedure for examining RQ2 depends on the outcome of RQ1. That is, if we identify multiple trajectory classes in the initial GMMs, gender is added to the final GMM model as a covariate to determine the likelihood of each gender being a member of the different developmental groups. If a single-profile solution best describes the data, a multigroup latent growth model is estimated to investigate whether the developmental trajectories (i.e., initial level and slope) and the predictions of the background variables differ between boys and girls. The Wald chi-square test will be used to test if significant growth parameters differ across genders (Wang & Wang, 2020). The detailed descriptions and steps regarding different analyses are given together with the results.

3. Results

3.1. Descriptive statistics and correlations

The descriptive statistics, reliabilities as McDonald's omega, and correlations between the variables are given in Table 1. All measures showed acceptable reliability, $\omega > 0.70$ (Tavakol & Dennick, 2011). The values of kurtosis in arithmetic t1 and t2 were fairly high, due to the fact that many children scored rather low and were not yet fluent in their skills, however, the kurtosis value diminished over the time. The rank-order stability of arithmetic fluency was rather high, with correlations between measurement points ranging from 0.68 to 0.87. Both domain-specific skills had moderate positive relations with arithmetic fluency; symbolic magnitude processing ranging between $r = 0.53$–0.57 and number sequence skills between $r = 0.55$–0.65. For the domain-general skills, the central executive and RAN correlated moderately with arithmetic fluency ($r = 0.37$–0.48), while non-verbal reasoning and attention showed only small positive associations with arithmetic fluency ($r = 0.16$–0.33). The correlations between all covariates were statistically significant and small to moderate ($r = 0.23$–0.50), and thus we did not expect any multicollinearity issues in later analyses.

3.2. The development of arithmetic fluency

As our measurement points were unevenly spaced (i.e., having a longer gap between t2 and t3 because of missing one data collection due to Covid-19), we first ran two separate univariate latent growth curve models (LGCM) for arithmetic fluency performance over time. In the first model, following the actual gaps between measurements, we fixed t1 and t2 at zero and one, and t3 and t4 at three and four (Preacher et al., 2008). However, this model did not describe the data well, χ^2 (5) = 67.486, $p < .000$; RMSEA = 0.218 (90 % CI = 0.174–0.266); CFI = 0.920; SRMR = 0.050. Therefore, we ran a latent basis model by fixing t1 at zero and t4 at one, and with t2 and t3 allowed to be freely estimated. This model showed a good fit to the data, χ^2 (3) = 1.561, $p = .668$; RMSEA = 0.000 (90 % CI =0.000–0.081); CFI = 1.000; SRMR = 0.008, and was hence used in further analyses. There was a significant increase in the fluency development over time (M slope = 25.71, $p < .001$), as well as individual differences (i.e., variance) shown in the initial level ($s^2 = 50.35$, $p < .001$) and slope as well ($s^2 = 115.41$, $p < .001$).

As the results from the univariate LGCM showed individual differences in arithmetic development, evidenced by statistically significant ($p < .001$) variances in both the intercept and slope, we ran a series of growth mixture models (GMM) with 2–5 latent classes in order to identify possible distinct developmental trajectories. Based on an overall evaluation of the competing models, we found the one group solution (i.e., one developmental trajectory) to be most appropriate. Although the statistical indicators suggested two or three groups (Table 2),[1] we considered these alternatives unfeasible due to the small number of children in the additional groups. In the 2-group model, the smaller group ($n = 19$) was a higher performing group, and in the 3-group model, an additional group ($n = 17$) with steeper growth was included. In our view, the model with an overall developmental trajectory (i.e., the 1-group solution) allowed for a more coherent and reliable further investigation to answer our remaining research questions. The descriptive statistics together with figures of the developmental trajectories for the 2- and 3-group models are provided in the Supplementary materials.

Next, we estimated a multigroup model with gender as a grouping variable to examine the similarity of change over time between boys and girls. The model showed a good fit to the data, χ^2 (8) = 8.218, $p = .412$; RMSEA = 0.014 (90 % CI =0.000–0.104); CFI = 1.000; SRMR = 0.030. As shown in Fig. 1, the developmental trajectories were similar for boys and girls. Based on Wald's test, the means of initial levels ($M_{\text{boys}} = 10.75$; $M_{\text{girls}} = 9.48$; χ^2 (1) = 1.823, $p = .177$) and slopes ($M_{\text{boys}} = 24.66$; $M_{\text{girls}} = 27.04$; χ^2 (1) = 2.599, $p = .107$) were not statistically different, thus indicating gender invariant development in arithmetic fluency.

3.3. Predictors of arithmetic development by gender

Next, we added predictors to the model. First, we specified the predictions to be invariant across gender, and checked whether the modification indices suggested some predictions to be different for boys and girls. The model fit to the data well, χ^2 (56) = 69.086, $p = .113$; RMSEA = 0.042 (90 % CI =0.000–0.072); CFI = 0.988; SRMR = 0.053. However, modification indices suggested two effects from the covariates to be non-invariant, mother's educational level on the intercept and RAN on the slope. Thus, in the next model, these two effects were estimated freely for boys and girls, with a slightly improved fit to the data, χ^2 (54) = 57.937, $p = .332$; RMSEA = 0.024 (90 % CI =0.000–0.061); CFI = 0.996; SRMR = 0.043. Based on the Wald test, both the prediction of mother's educational level on the initial level, χ^2 (1) = 4.84, $p = .028$, and the prediction of RAN on the rate of change, χ^2 (1) = 7.50, $p = .006$, were indeed different for boys and girls. The effect of mother's education on the onset of arithmetic development was significant for boys ($\beta = 0.17$, $p = .001$) but not for girls ($\beta = 0.01$, $p = .920$), while the effect of RAN on the slope of arithmetic development was significant for girls ($\beta = 0.33$, $p \leq 0.001$) but not for boys ($\beta = 0.02$, $p = .870$).

As to the gender invariant effects (Table 3), symbolic magnitude processing, number sequence skills, and working memory were found to predict the initial level of arithmetic fluency, whereas symbolic magnitude processing and number sequence skills were found to be predictors of growth in arithmetic fluency. On this basis, the domain-specific skills were found to be better predictors of arithmetic fluency development than domain-general skills, except for working memory,

---

[1] Freeing within class variances and/or covariances did not improve the model fit.

Table 1
Correlations and Descriptive Statistics.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Arithmetic T1	–									
2. Arithmetic T2	0.77***	–								
3. Arithmetic T3	0.71***	0.81***	–							
4. Arithmetic T4	0.68***	0.79***	0.87***	–						
5. SMP	0.53***	0.54***	0.57***	0.56***	–					
6. Number sequences	0.65***	0.55***	0.57***	0.60***	0.50***	–				
7. WM: central executive	0.44***	0.37***	0.41***	0.43***	0.28***	0.46***	–			
8. RAN	0.40***	0.41***	0.48***	0.46***	0.48***	0.36***	0.37***	–		
9. Nonverbal reasoning	0.28***	0.22***	0.29***	0.33***	0.30***	0.39***	0.39***	0.28***	–	
10. Attention	0.25***	0.16**	0.24***	0.24**	0.44***	0.23***	0.23**	0.24***	0.27***	–
M	10.17	18.89	28.28	35.89	17.28	19.67	6.06	9.78	23.78	64.72
SD	7.77	10.53	13.37	16.44	4.18	6.24	1.80	2.50	5.25	19.24
Skewness	1.46	1.22	0.72	0.51	0.04	−0.76	−0.11	0.10	−0.35	−0.45
Kurtosis	3.23	2.76	0.49	0.26	−0.22	−0.09	0.69	0.89	−0.25	0.95
McDonald's ω	0.899	0.959	0.977	0.982	0.880	0.869	0.730	0.863	0.849	–

Note. SNP = Symbolic Magnitude Processing; WM = Working Memory; RAN = Rapid Automatized Naming. T1–T4 = Time 1 – Time 4.
*** $p < .001$; ** $p < .01$.

Table 2
Fit Indices for Distinct Arithmetic Trajectory Groups.

N of profiles	BIC	aBIC	AIC	Entropy	*p* LMR	*p* VLMR	n class 1 (ALCP)	n class 2 (ALCP)	n class 3 (ALCP)	n class 4 (ALCP)	n class 5 (ALCP)
2	7185.413	7141.027	7135.456	0.939	0.0056	0.0044	19 (0.990)	243 (0.873)	–	–	–
3	7184.001	7130.103	7123.339	0.869	0.0069	0.0050	21 (0.853)	17 (0.805)	224 (0.964)	–	–
4	7186.898	7123.489	7115.531	0.841	0.2242	0.2090	7 (0.942)	34 (0.831)	17 (0.870)	204 (0.929)	–
5	7197.137	7124.217	7115.065	0.818	0.5281	0.5187	2 (0.995)	11 (0.944)	16 (0.830)	192 (0.908)	14 (0.795)

Note. BIC = Bayesian Information Criterion; aBIC = Adjusted Bayesian Information Criterion; AIC = Akaike's Information Criterion; LMR = Lo-Mendell-Rubin Adjusted Likelihood Ratio Test; VLMR = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test; ALCP = Average Latent Class Probabilities for Most Likely Latent Class Membership by Latent Class.

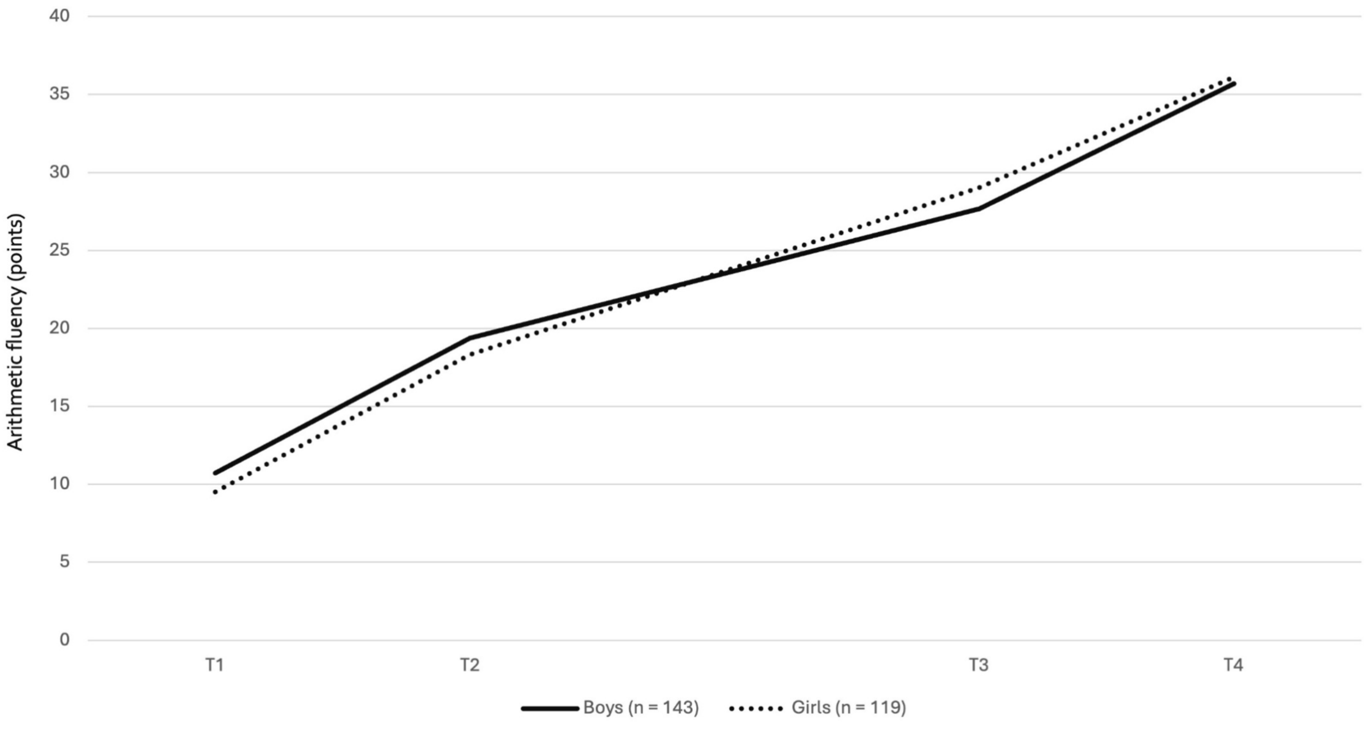


Fig. 1. The development of arithmetic fluency from first to third grade by gender.

Table 3

Unstandardized effects on the initial level and change in arithmetic fluency by gender.

Predictor	Arithmetic fluency girls ($n = 119$)				Arithmetic fluency boys ($n = 143$)			
	Intercept		Slope		Intercept		Slope	
	B (SE)	*z*	*B (SE)*	*z*	*B (SE)*	*z*	*B (SE)*	*z*
SMP	0.451 (0.111)	4.060***	0.741 (0.235)	3.150**	0.451 (0.111)	4.060***	0.741 (0.235)	3.150**
Number sequences	0.470 (0.070)	6.684***	0.343 (0.129)	2.655**	0.470 (0.070)	6.684***	0.343 (0.129)	2.655**
WM: central executive	0.789 (0.272)	2.901**	0.271 (0.389)	0.696	0.789 (0.272)	2.901**	0.271 (0.389)	0.696
RAN	0.349 (0.196)	1.783	1.490 (0.480)	3.106***	0.349 (0.196)	1.783	0.071 (0.431)	0.164
Nonverbal reasoning	−0.096 (0.078)	−1.237	0.136 (0.152)	0.897	−0.096 (0.078)	−1.237	0.136 (0.152)	0.897
Attention	−0.001 (0.021)	−0.029	−0.036 (0.039)	−0.919	−0.001 (0.021)	−0.029	−0.036 (0.039)	−0.919
Age	0.008 (0.106)	0.076	0.298 (0.226)	1.315	0.008 (0.106)	0.076	0.298 (0.226)	1.315
Mother's education	0.057 (0.568)	0.100	−1.054 (0.889)	−1.185	1.870 (0.611)	3.058**	−1.054 (0.889)	−1.185
R^2	0.542		0.409		0.583		0.242	

Note. SNP = Symbolic Magnitude Processing; WM = Working Memory; RAN = Rapid Automatized Naming.
*** $p < .001$; ** $p < .01$; * $p < .05$.

which also predicted the initial level.

4. Discussion

We aimed to extend current knowledge about children's development of arithmetic fluency from first to third grade, to identify possible distinct trajectories in this development, and to determine which domain-specific, domain-general, and demographic factors predict this development. Furthermore, we were interested in whether gender moderates the development of arithmetic fluency or the relations between the predictors and arithmetic fluency. Contrary to our hypothesis H1.1, we found that one trajectory group best described the development of arithmetic fluency. Girls and boys developed similarly in their arithmetic fluency over time, hence confirming H1.2. However, gender differences were observed in the predictors of arithmetic fluency development. While symbolic magnitude processing, number sequence skills and working memory predicted the first-grade arithmetic performance similarly for boys and girls, mother's educational level predicted the first-grade arithmetic performance only for boys. Regarding growth, symbolic magnitude processing and number sequence skills were significant predictors regardless of gender, whereas RAN predicted girls' arithmetic improvement over time.

As expected, children's arithmetic fluency improved over time from first to third grade. At the end of first grade, many children use counting-based procedural strategies, rather than retrieval to produce the answers (Koponen et al., 2007). Gradually, with practice, they begin to be able to retrieve the facts quickly from long-term memory, which shows up as improved fluency. Until recently, many prior studies have used artificial cut-off criteria for grouping of children into different performance groups. Here, we applied a data-driven, person-centered approach to find trajectories of arithmetic fluency that differ in initial level and/or growth. One trajectory group was found to best describe the data when the criteria used to determine the number of classes were followed, even though individual differences were observed in both the children's initial level and growth. The 2-group and 3-group solutions included rather small groups in addition to the mainstream group, and surprisingly, only one higher performing group (in a 2-group solution) or one group with steeper growth (in a 3-group solution). A few other studies using a person-centered approach have found different profiles (Bakker et al., 2023; Scalise et al., 2021; Zhang et al., 2020). However, Bakker et al. (2023) and Scalise et al. (2021) used a more comprehensive mathematics test as a measure of performance, which may have captured bigger variance in performance than the arithmetic fluency measure used in this study. Zhang et al. (2020) used a similar measure of arithmetic fluency and found distinct developmental trajectories from first to fourth grade. These trajectories showed rather similar development and differed mainly in fourth grade performance, which may partly explain the different results from our study, as we only measured arithmetic skills up to third grade. In addition, their sample size was larger, which may make it easier to find subgroups.

In this age group, gender differences in mathematics performance seem to be rather non-existent (Hyde et al., 2008; Lachance & Mazzocco, 2006), as our study also showed. Boys and girls did not differ in either their initial level or growth in arithmetic fluency. Apart from a few studies in this age group that have focused on arithmetic skills and found either boys (Xu et al., 2021) or girls (Lachance & Mazzocco, 2006) performing better, gender differences seem to emerge only in later grades (Martens et al., 2011; Räsänen et al., 2021). Therefore, it would be interesting to conduct a similar study in later grades of primary school, and even including all four different arithmetic operations. While gender should not be a guiding principle in mathematics education, it is important to be aware of possible gender differences. In particular, concerns have been raised about how to motivate girls to pursue STEM studies and careers in which they are underrepresented (The World Economic Forum, 2023). Making the learning of mathematics interesting, appreciated and useful for all, using a variety of teaching methods, should start at the beginning of school and continue throughout the school years.

This study is one of the first to simultaneously include both symbolic magnitude processing and number sequence skills as domain-specific predictors of arithmetic fluency. We found that they predicted both initial performance and growth, regardless of gender. Both skills tap number processing, but in symbolic comparison, the focus is on cardinal (magnitude) information, whereas in number sequence performance, the focus is on ordinal (order) information (Devlin et al., 2022). A previous study (Vanbinst et al., 2016) suggested that the role of symbolic magnitude processing is as important for arithmetic as phonological awareness is for reading. Our results support this, especially since symbolic magnitude processing was a predictor of growth in arithmetic fluency. Research investigating the link and common mechanisms between symbolic magnitude processing and arithmetic fluency is still in its infancy, and more longitudinal studies are needed. Further, reliable and brief tests are available to assess children's symbolic magnitude processing skills, but whether and how teaching could take into account individual differences in symbolic magnitude processing skills is still an open question, and thus intervention studies are also urgently needed in the future.

Previously, Banfi et al. (2024) found magnitude processing skills (assessed as dots and digit comparison tasks in grade 1) to be a better predictor of arithmetic fluency in grade 3 than number sequence skills (assessed only as counting forwards from 1 for 1 min in grade 1). In their study, the measures of magnitude processing and number sequences differed slightly from ours, and only predictions at different time points (grades 1–3) were examined, not growth. Our findings also support the role of number sequence skills as a unique predictor of arithmetic fluency (Koponen et al., 2019; Träff et al., 2023; Zhang et al., 2017). In contrast to Träff et al. (2023), we found that number sequence skills predicted not only initial performance but also development. This finding underscores that number sequence skills are critical skills to practice and master early in school. Children with fluent number sequence skills, such as being able to start counting from different points in the number sequence (e.g., starting at 8 or 27, counting forward or backward) and being able to recite the number sequence correctly, may be better able to use their accurate and efficient counting strategy to solve arithmetic problems. This, in turn, could lead to a faster development of arithmetic facts compared to children whose number sequencing skills are weaker (Gilmore & Batchelor, 2021). The mathematics curricula and teaching materials for pre-primary education and early grades include a variety of counting tasks, both number sequence and object counting tasks, to guide teachers in implementing counting activities in their classrooms, taking into account individual differences in children's development.

Working memory, more specifically CE, was found to be associated with arithmetic fluency in first grade, but not with the growth of arithmetic fluency over time. This finding is similar to Wei, Guo, et al. (2018), but contrasts with the results of some previous studies, which found that CE also predicted the development of arithmetic fluency (Hoff et al., 2023). Furthermore, our results are consistent with those of Barrouillet and Lépine (2005) and Geary et al. (2012), who showed that the children in the early grades with higher CE capacity used fact retrieval in addition problems more frequently than their peers with lower CE capacity. In other words, as CE capacity increases, the child will be more likely to retrieve facts from long-term memory and thus show better fluency. According to our results, CE ability in first grade does not seem to affect fact learning (i.e., growth) in the early school years, a finding similar to Hoff et al.'s (2023) first graders. The role of CE needs further investigation, especially from a longitudinal perspective.

Overall, we found that boys and girls were more similar than different in cognitive functions (Hyde, 2016). However, gender moderated the relations between RAN and growth in arithmetic fluency and between mother's educational level and intercept of arithmetic fluency. RAN predicted the growth in arithmetic fluency for girls but not for boys. In prior research, Lachance and Mazzocco (2006) found girls to outperform boys on the Colors subtest of RAN, but they had a cross-sectional design. Unfortunately, other previous studies that have included a mathematics component have not reported on the role of gender in RAN (Hoff et al., 2023; Koponen et al., 2017, 2016; Zhang et al., 2020). Thus, RAN and its relation to numeracy skills, accompanied by a gender perspective, needs to be further explored in future studies. Since the RAN test requires children to express their answers verbally, it would be interesting to see if expressive language skills interact here and could possibly explain the gender difference we found (e.g., Eriksson et al., 2012).

Mother's educational level predicted the initial level of boys' arithmetic fluency. In prior studies conducted in Nordic countries, parents' educational level has had little impact on children's early mathematics learning (Aunio & Niemivirta, 2010). Since we did not find an influence of the mother's educational level on the growth in arithmetic skills, other factors, in this case domain-specific mathematical skills, seem to override this effect. To better capture and understand why mother's educational level was connected to boys' arithmetic performance in the beginning of schooling, future studies could further explore the role of parents as part of children's home numeracy environment and its relationship to the development of arithmetic fluency skills (Khanolainen et al., 2023; Mutaf-Yıldız et al., 2020).

4.1. Limitations and future directions

Our study has some limitations that should be addressed when interpreting the results and in future studies. First, due to our study design, that is, the predictor variables were measured only once, it is obvious that definite conclusions about causality cannot be drawn. For example, it is possible that improved arithmetic fluency would also affect the development of number sequence and symbolic magnitude processing skills. Second, our sample size, although sufficient for a longitudinal study and for chosen analyses, may have restricted finding big enough distinct trajectory groups (Ferguson et al., 2020).

In our study, we focused on arithmetic fluency, which is one of the fundamental skills to be learnt in primary school. Other similar studies could be conducted in the future, focusing on overall mathematics achievement or profiling a range of other mathematical subskills (Bakker et al., 2023). The domain-specific and domain-general factors in our study were selected based on prior literature. Most of the skills have proven to be associated with mathematics performance, while some have been less studied, such as sustained attention. We recognize that our study did not measure some important predictors that have been shown to be related to mathematics performance, such as various language skills (Bakker et al., 2023) or the role of home environment (DePascale et al., 2023). In our study, we did not separate arithmetic skills into addition and subtraction, nor did we separate number sequences into forward and backward sequences. In future studies, it may be interesting to examine these constructs separately to see if, for example, forward number sequence skills are more strongly related to addition fluency than subtraction fluency. With a larger sample, it would be possible to include even more predictors to get a more comprehensive view of the simultaneous effects of both domain-specific and domain-general skills, as well as different background characteristics of children, on arithmetic development.

4.2. Conclusions

As expected, children's arithmetic skills improved over time from first to third grade, and there were individual differences in both initial levels in first grade and in growth. Contrary to our expectations, we captured only one trajectory group that best described children's development of arithmetic fluency. Symbolic magnitude processing, number sequence skills, and working memory predicted first-grade arithmetic performance for both boys and girls, whereas mother's education level was associated with initial arithmetic fluency performance only for boys. Looking at growth in arithmetic fluency, symbolic magnitude processing and number sequence skills were significant predictors regardless of gender, whereas RAN predicted only girls' improvement over time. To conclude, the domain-specific factors, symbolic magnitude processing and number sequence skills, were related to growth in arithmetic fluency (while the domain-general factors were not), and appear to be promising skills to focus on when teaching mathematics at the beginning of schooling.

Code availability

Upon request from the correspondent author.

Consent for participation

Written consent from parents and teachers.

Educational relevance and implications statement

One overall developmental trajectory best described children's development of arithmetic fluency from first to third grade. No gender differences were found. The domain-specific factors, symbolic magnitude processing, and number sequence skills, were related to both initial level and growth in arithmetic fluency, whereas the domain-general factors were related only to initial level of arithmetic fluency. This finding emphasizes that symbolic magnitude processing and number sequence skills are critical skills to practice and master at the beginning of schooling.

CRediT authorship contribution statement

Riikka Mononen: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Johan Korhonen:** Writing – original draft, Methodology, Formal analysis, Conceptualization. **Karoline Hægeland:** Writing – review & editing, Investigation, Conceptualization. **Matin Younesi:** Writing – review & editing, Investigation, Conceptualization. **Silke M. Göbel:** Writing – review & editing, Conceptualization. **Markku Niemivirta:** Writing – original draft, Methodology, Formal analysis, Conceptualization.

Ethics approval

Norwegian Centre for Research Data.

Consent for publication

Written consent from parents and teachers.

Funding

This work was supported by the Norwegian Research Council [Grant number: 283396], for Riikka Mononen and partially supported by funding of the Economic and Social Research Council [ES/N014677/1, ES/W002914/1] to Silke Göbel.

Declaration of competing interest

None.

Acknowledgements

We thank all participating students and their teachers, as well as research assistants who took part in the data collections.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lindif.2024.102585.

Data availability

Upon request from the correspondent author.

References

Allen, K., Giofrè, D., Higgins, S., & Adams, J. (2020). Working memory predictors of written mathematics in 7- to 8-year-old children. *Quarterly Journal of Experimental Psychology, 73*(2), 239–248. https://doi.org/10.1177/1747021819871243

Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *British Journal of Educational Psychology, 78*(2), 181–203. https://doi.org/10.1348/000709907X209854

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20*(5), 427–435. https://doi.org/10.1016/j.lindif.2010.06.003

Aunio, P., & Räsänen, P. (2016). Core numerical skills for learning mathematics in children aged five to eight years – A working model for educators. *European Early Childhood Education Research Journal, 24*(5), 684–704. https://doi.org/10.1080/1350293X.2014.996424

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*(4), 699–713. https://doi.org/10.1037/0022-0663.96.4.699

Bakker, M., Torbeyns, J., Verschaffel, L., & De Smedt, B. (2023). Longitudinal pathways of numerical abilities in preschool: Cognitive and environmental correlates and relation to primary school mathematics achievement. *Developmental Psychology, 59*(3), 442–459. https://doi.org/10.1037/dev0001491

Banfi, C., Jöbstl, V., Göbel, S. M., & Landerl, K. (2024). Longitudinal predictors of reading and arithmetic at different attainment levels. *Frontiers in Education, 9*. https://doi.org/10.3389/feduc.2024.1335957

Barrouillet, P., & Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology, 91*(3), 183–204. https://doi.org/10.1016/j.jecp.2005.03.002

Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP test). *Behavior Research Methods, 49*(4), Article 4. https://doi.org/10.3758/s13428-016-0792-3

Cai, D., Zhang, L., Li, Y., Wei, W., & Georgiou, G. K. (2018). The role of approximate number system in different mathematics skills across grades. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01733

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*(1), 61–75. https://doi.org/10.1016/j.lindif.2007.04.005

Chen, E. H., & Bailey, D. H. (2021). Dual-task studies of working memory and arithmetic performance: A meta-analysis. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 47*(2), 220–233. https://doi.org/10.1037/xlm0000822

Cohen, R. A. (2011). Sustained attention. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 2440–2443). Springer. https://doi.org/10.1007/978-0-387-79948-3_1334.

Cvencek, D., Brečić, R., Gaćeša, D., & Meltzoff, A. N. (2021). Development of math attitudes and math self-concepts: Gender differences, implicit–explicit dissociations, and relations to math achievement. *Child Development, 92*(5), e940–e956. https://doi.org/10.1111/cdev.13523

De Smedt, B. (2022). Individual differences in mathematical cognition: A Bert's eye view. *Current Opinion in Behavioral Sciences, 46*, Article 101175. https://doi.org/10.1016/j.cobeha.2022.101175

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

De Smedt, B., Noël, M.-P., Gilmore, C., & 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 in Neuroscience and Education, 2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103*(4), 469–479. https://doi.org/10.1016/j.jecp.2009.01.010

DePascale, M., Jaeggi, S. M., & Ramani, G. B. (2023). The influence of home environmental factors on kindergarten children's addition strategy use. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1027431

Devlin, D., Moeller, K., Reynvoet, B., & Sella, F. (2022). A critical review of number order judgements and arithmetic: What do order verification tasks actually measure? *Cognitive Development, 64*, Article 101262. https://doi.org/10.1016/j.cogdev.2022.101262

Efrat, B., & Orna, T. (2022). Sex differences in the sustained attention of elementary school children. *BMC Psychology, 10*(1), 307. https://doi.org/10.1186/s40359-022-01007-z

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 13*(2), 190–193. https://doi.org/10.1177/1745691617720478

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., … Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *The British Journal of Developmental Psychology, 30*(Pt 2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Ferguson, S. L., Moore, G., W, E., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development, 44*(5), 458–468. https://doi.org/10.1177/0165025419881721

Finke, S., Vogel, S. E., Freudenthaler, H. H., Banfi, C., Steiner, A. F., Kemény, F., … Landerl, K. (2022). Developmental trajectories of symbolic magnitude and order processing and their relation with arithmetic development. *Cognitive Development, 64*, Article 101266. https://doi.org/10.1016/j.cogdev.2022.101266

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44. https://doi.org/10.1016/j.edurev.2013.05.003

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., … Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29–43.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., … Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*(6), 1731–1746. https://doi.org/10.1037/a0020662

Fuson, K., & Grouws, D. A. (1992). Research on whole number addition and subtraction. In *Handbook of research on mathematics teaching and learning* (pp. 243–275). NCTM.

Geary, D. C. (2011a). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552. https://doi.org/10.1037/a00 25510

Geary, D. C. (2011b). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics: JDBP, 32*(3), 250–263. https://doi.org/10.1097/DBP.0b0 13e318209edef

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & Catherine DeSoto, M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*(2), 121–151. https://doi.org/10.1016/j.jecp.2004.03.002

Geary, D. C., Hoard, M. K., & Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. *Journal of Experimental Child Psychology, 113*(1), 49–65. https://doi.org/10.1016/j.jecp.2012.03.003

Gilmore, C., & Batchelor, S. (2021). Verbal count sequence knowledge underpins numeral order processing in children. *Acta Psychologica, 216*, Article 103294. https://doi.org/10.1016/j.actpsy. 2021.103294

Gray, S. A., Rogers, M., Martinussen, R., & Tannock, R. (2015). Longitudinal relations among inattention, working memory, and academic achievement: Testing mediation and the moderating role of gender. *PeerJ, 3*, Article e939. https://doi.org/10.7717/peerj.939

Habermann, S., Donlan, C., Göbel, S. M., & Hulme, C. (2020). The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development. *Journal of Experimental Child Psychology, 193*, Article 104794. https://doi.org/10.1016/j.jecp.2019.104794

Hasson, R., & Fine, J. G. (2012). Gender differences among children with ADHD on continuous performance tests: A meta-analytic review. *Journal of Attention Disorders, 16*(3), 190–198. https://doi.org/10.1177/1087054711427398

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences, 66*, 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Hoff, D., Amland, T., Melby-Lervåg, M., Lervåg, A., & Protopapas, A. (2023). Early rapid naming longitudinally predicts shared variance in reading and arithmetic fluency. *Journal of Experimental Child Psychology, 231*, Article 105656. https://doi.org/10.1016/j.jecp.2023.105656

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Hornung, C., Martin, R., & Fayol, M. (2017). General and specific contributions of RAN to reading and arithmetic fluency in first graders: A longitudinal latent variable approach. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01746

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyde, J. S. (2016). Sex and cognition: Gender and cognitive functions. *Current Opinion in Neurobiology, 38*, 53–56. https://doi.org/10.1016/j.conb.2016.02.007

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*(5888), 494–495. https://doi.org/10.1126/science.1160364

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*(2), 103–119. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2791329/.

Khanolainen, D., Koponen, T., Eklund, K., Gerike, G., Psyridou, M., Lerkkanen, M.-K., Aro, M., & Torppa, M. (2023). Parental influences on the development of single and co-occurring difficulties in reading and arithmetic fluency. *Learning and Individual Differences, 105*, Article 102321. https://doi.org/10.1016/j.lindif.2023.102321

Klausen, T., & Reikerås, E. (2016). *RegnefaktaprØven*. Lesesenteret, University of Stavenger.

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J.-E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skill. *Journal of Experimental Child Psychology, 97*(3), 220–241. https://doi.org/10.1016/j.jecp.2007.03.001

Koponen, T., Aunola, K., & Nurmi, J.-E. (2019). Verbal counting skill predicts later math performance and difficulties in middle school. *Contemporary Educational Psychology, 59*, Article 101803. https://doi.org/10.1016/j.cedpsych.2019.101803

Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology, 109*(7), 977–992. https://doi.org/10.1037/edu0000182

Koponen, T., Salmi, P., Torppa, M., Eklund, K., Aro, T., Aro, M., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2016). Counting and rapid naming predict the fluency of arithmetic and reading skills. *Contemporary Educational Psychology, 44–45*, 83–94. https://doi.org/10.1016/j.cedpsych.2016.02.004

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education, 23*(1), 77–94. https://doi.org/10.1007/BF03173141

Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*(3), 195–216. https://doi.org/10.1016/j.lindif.2005.12.001

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395–438. https://doi.org/10.1016/j.cognition.2006.10.005

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*(5), 481–498. https://doi.org/10.1016/j.intell.2004.06.008

Martens, R., Hurks, P. P. M., Meijs, C., Wassenberg, R., & Jolles, J. (2011). Sex differences in arithmetical performance scores: Central tendency and variability. *Learning and Individual Differences, 21*(5), 549–554. https://doi.org/10.1016/j.lindif.2011.06.003

Mutaf-Yıldız, B., Sasanguie, D., De Smedt, B., & Reynvoet, B. (2020). Probing the relationship between home numeracy and children's mathematical skills: A systematic review. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.02074

Muthén, L. K., & Muthén, B. O. (1998). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

Orbach, L., & Fritz, A. (2022). Patterns of attention and anxiety in predicting arithmetic fluency among school-aged children. *Brain Sciences, 12*(3), 376. https://doi.org/10.3390/brainsci12030376

Pelegrina, S., Lechuga, M. T., García-Madruga, J. A., Elosúa, M. R., Macizo, P., Carreiras, M., … Bajo, M. T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01544

Pina, V., Fuentes, L. J., Castillo, A., & Diamantopoulou, S. (2014). Disentangling the effects of working memory, language, parental education, and non-verbal intelligence on children's mathematical abilities. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00415

Preacher, K. J., Wichman, A. J., MacCallum, R. C., & Briggs, N. E. (2008). *Latent Growth Curve Modeling*. Sage.

Pulkkinen, J., Eklund, K., Koponen, T., Heikkilä, R., Georgiou, G., Salminen, J., van Daal, V., & Aro, M. (2022). Cognitive skills, self-beliefs and task interest in children with low reading and/or arithmetic fluency. *Learning and Individual Differences, 97*, Article 102160. https://doi.org/10.1016/j.lindif.2022.102160

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., … 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. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.683672

Raven, J. C., Court, J. H., & Raven, J. (1990). *Manual for Raven's progressive matrices and vocabulary scales: Section 2*. Coloured progressive matrices: Oxford University Press.

Rawlings, A. M., Niemivirta, M., Korhonen, J., Lindskog, M., Tuominen, H., & Mononen, R. (2023). Achievement emotions and arithmetic fluency - Development and parallel processes during the early school years. *Learning and Instruction, 86*, Article 101776. https://doi.org/10.1016/j.learninstruc.2023.101776

Roid, G. H., Miller, M. J., Pomplun, M., & Koch, C. (2013). *Leiter international performance scale* (3rd ed.). Stoelting.

Rosselli, M., Ardila, A., Matute, E., & Inozemtseva, O. (2009). Gender differences and cognitive correlates of mathematical skills in school-aged children. *Child Neuropsychology, 15*(3), 216–231. https://doi.org/10.1080/09297040802195205

Salminen, J., & Koponen, T. (2011). Oppimisen arviointi: Matematiikan oppimisen seurannan välineet. In *Käsikirja [LukiMat–assessment of learning: Progress monitoring of mathematics learning. Manual]*. Niilo Mäki Institute.

Scalise, N. R., Daubert, E. N., & Ramani, G. B. (2021). When one size does not fit all: A latent profile analysis of low-income preschoolers' math skills. *Journal of Experimental Child Psychology, 209*, Article 105156. https://doi.org/10.1016/j.jecp.2021.105156

Scalise, N. R., & Ramani, G. B. (2021). Symbolic magnitude understanding predicts preschoolers' later addition skills. *Journal of Cognition and Development, 22*(2), 185–202. https://doi.org/10.1080/15248372.2021.1888732

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2). https://www.psycharchives.org/en/item/1a8dea48-0285-4dac-a612-9dc0ff2532f6.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical evaluation of language fundamentals–Fourth edition (CELF-4). Norwegian version*. Pearson.

Siegler, R., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do?. In *C. Sophian, origins of cognitive skills* (pp. 229–293). Lawrence Erlbaum Associates.

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173–181. https://doi.org/10.1016/j.lindif.2018.12.005

Sussman, J., & Tasso, A. F. (2013). The Mesulam continuous performance test (M-CPT): Age-related and gender differences in the sustained attention of elementary school children. *The New School Psychology Bulletin, 10*(2), 1–13. https://www.nspb.net/index.php/nspb/article/view/199.

Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Developmental Science, 17*(4), 506–524. https://doi.org/10.1111/desc.12144

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

The World Economic Forum. (2023). Global gender gap report 2023. https://www.weforum.org/publications/global-gender-gap-report-2023/.

Toivainen, T., Papageorgiou, K. A., Tosto, M. G., & Kovas, Y. (2017). Sex differences in non-verbal and verbal abilities in childhood and adolescence. *Intelligence, 64*, 81–88. https://doi.org/10.1016/j.intell.2017.07.007

Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *Journal of Experimental Child Psychology, 116*(2), 139–156. https://doi.org/10.1016/j.jecp.2013.04.007

Träff, U., Olsson, L., Östergren, R., & Skagerlund, K. (2020). Development of early domain-specific and domain-general cognitive precursors of high and low math achievers in grade 6. *Child Neuropsychology, 26*(8), 1065–1090. https://doi.org/10.1080/09297049.2020.1739259

Träff, U., Skagerlund, K., Östergren, R., & Skagenholt, M. (2023). The importance of domain-specific number abilities and domain-general cognitive abilities for early arithmetic achievement and development. *British Journal of Educational Psychology, 93*(3), 825–841. https://doi.org/10.1111/bjep.12599

Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic numerical magnitude processing is as important to arithmetic as phonological awareness is to reading. *PLoS One, 11*(3), Article e0151045. https://doi.org/10.1371/journal.pone.0151045

Vanbinst, K., Ceulemans, E., Ghesquière, P., & De Smedt, B. (2015). Profiles of children's arithmetic fact development: A model-based clustering approach. *Journal of Experimental Child Psychology, 133*, 29–46. https://doi.org/10.1016/j.jecp.2015.01.003

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2014). Arithmetic strategy development and its domain-specific and domain-general cognitive correlates: A longitudinal study in children with persistent mathematical learning difficulties. *Research in Developmental Disabilities, 35*(11), 3001–3013. https://doi.org/10.1016/j.ridd.2014.06.023

Venker, C. C., Goodwin, J. L., Roe, D. J., Kaemingk, K. L., Mulvaney, S., & Quan, S. F. (2007). Normative psychomotor vigilance task performance in children ages 6 to 11—The Tucson Children's assessment of sleep apnea (TuCASA). *Sleep and Breathing, 11*(4), 217–224. https://doi.org/10.1007/s11325-007-0103-4

Vuontela, V., Steenari, M.-R., Carlson, S., Koivisto, J., Fjällberg, M., & Aronen, E. T. (2003). Audiospatial and visuospatial working memory in 6–13 year old school children. *Learning & Memory, 10*(1), 74–81. https://doi.org/10.1101/lm.53503

Wang, J., & Wang, X. (2020). *Structural equation modeling: Applications using Mplus* (2nd ed.). Wiley.

Wang, L. (2020). Mediation relationships among gender, spatial ability, math anxiety, and math achievement. *Educational Psychology Review, 32*(1), Article 1. https://doi.org/10.1007/s10648-019-09487-z

Wechsler, D. (2017). *WISC-V Wechsler intelligence scale for children – Fifth edition* (Norwegian version).

Wei, W., Deng, W., Chen, C., He, J., Qin, J., & Kovas, Y. (2018). A mechanistic study of the association between symbolic approximate arithmetic performance and basic number magnitude processing based on task difficulty. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01551

Wei, W., Guo, L., Georgiou, G. K., Tavouktsoglou, A., & Deng, C. (2018). Different subcomponents of executive functioning predict different growth parameters in mathematics: Evidence from a 4-year longitudinal study with Chinese children. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01037

West, G., Shanks, D. R., & Hulme, C. (2021). Sustained attention, not procedural learning, is a predictor of reading, language and arithmetic skills in children. *Scientific Studies of Reading, 25*(1), 47–63. https://doi.org/10.1080/10888438.2020.1750618

Widlund, A., Niemivirta, M., Tuominen, H., & Korhonen, J. (2024). Growth trajectories of self-concept and interest in mathematics and language – Individual differences and cross-domain relations. *Learning and Instruction, 91*, Article 101882. https://doi.org/10.1016/j.learninstruc.2024.101882

Xenidou-Dervou, I., Van Luit, J. E. H., Kroesbergen, E. H., Friso-van den Bos, I., Jonkman, L. M., van der Schoot, M., & van Lieshout, E. C. D. M. (2018). Cognitive predictors of children's development in mathematics achievement: A latent growth modeling approach. *Developmental Science, 21*(6), Article e12671. https://doi.org/10.1111/desc.12671

Xu, C., LeFevre, J.-A., Skwarchuk, S.-L., Di Lonardo Burr, S., Lafay, A., Wylie, J., … Simms, V. (2021). Individual differences in the development of children's arithmetic fluency from grades 2 to 3. *Developmental Psychology, 57*(7), 1067–1079. https://doi.org/10.1037/dev0001220

Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2017). Knowing, applying, and reasoning about arithmetic: Roles of domain-general and numerical skills in multiple domains of arithmetic learning. *Developmental Psychology, 53*(12), 2304–2318. https://doi.org/10.1037/dev0000432

Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2020). Early cognitive precursors of children's mathematics learning disability and persistent low achievement: A 5-year longitudinal study. *Child Development, 91*(1), 7–27. https://doi.org/10.1111/cdev.13123