Arithmetic Complexity in Multi-Digit Calculation: Paradigms, Attitudes, and Neural Correlates

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Xinru Yao
aus Jiangsu China

Tübingen
2026

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	08.05.2026
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Hans-Christoph Nuerk
2. Berichterstatter/-in:	Prof. Dr. Bert De Smedt

Table of Contents

List of figures and tables

Abstract (English)

Arithmetic skills are important in daily life, education, and career development. Yet laboratory research on arithmetic often employs decision paradigms and single-digit tasks, which raises the question of generalizability to production paradigms. Moreover, findings drawn from low math anxious samples may underestimate the difficulties experienced by high math anxious individuals, particularly in complex arithmetic tasks involving multi-digit numbers.

To deal with these issues, this dissertation focuses on carry/borrow effects in multi-digit addition and subtraction across paradigms, math attitudes, and neural processes in four preregistered studies. Studies 1 and 2 compared decision and production paradigms in two-digit arithmetic regarding cognition and emotion. It was found that the choice of paradigm affected the observed arithmetic effects, particularly accuracy, and that production paradigms elicited higher state math anxiety than decision paradigms, particularly in individuals with higher trait anxiety. Study 3 examined math attitudes across the lifespan, showing that math anxiety increases and math self-concept decreases with age, with a lower math self-concept being associated with higher arithmetic complexity effects. Study 4 used functional near-infrared spectroscopy (fNIRS) to explore how domain-general and domain-specific brain resources contribute to complexity in three-digit arithmetic. Results showed that fronto-parietal activation increased with arithmetic complexity and this increase was moderated by working memory capacity.

In sum, this dissertation shows that arithmetic with carry/borrow operations is not always equally difficult but also affected by task context (paradigm) as well as individual attitudes and working memory resources. This has implications for how arithmetic skills should be assessed and supported in educational and applied settings.

Zusammenfassung (German)

Rechenfähigkeiten sind wichtig für Alltag, Bildung und Beruf. In Laborstudien zum Rechnen werden jedoch häufig Entscheidungsparadigmen und einstellige Aufgaben verwendet, was zur Frage der Generalisierbarkeit auf Produktionsparadigmen führt. Zudem könnten Ergebnisse aus Stichproben mit geringer Mathematikangst die Probleme von Personen mit hoher Mathematikangst unterschätzen, insbesondere bei komplexen Rechenaufgaben mit mehrstelligen Zahlen.

Diese Dissertation untersucht daher Übertragseffekte in der Addition und Subtraktion mehrstelliger Zahlen über Paradigmen, Einstellungen zur Mathematik und neuronale Prozesse hinweg in vier präregistrierten Studien. In Studien 1 und 3 wurden Entscheidungs- und Produktionsparadigmen beim zweistelligen Rechnen hinsichtlich Kognition und Emotion verglichen. Es zeigte sich, dass die Wahl des Paradigmas die beobachteten arithmetischen Effekte, insbesondere die Genauigkeit, beeinflusste und dass Produktionsparadigmen eine höhere situative Mathematikangst hervorriefen als Entscheidungsparadigmen, insbesondere bei Personen mit höherer Trait-Angst. Studie 3 untersuchte die Einstellung zur Mathematik über die Lebensspanne und zeigte, dass die Mathematikangst im Laufe des Lebens zunimmt und das mathematische Selbstkonzept abnimmt, wobei ein geringeres Selbstkonzept in Mathematik mit einem höheren Übertragseffekt assoziiert war. In Studie 4 wurde die funktionelle Nahinfrarotspektroskopie (fNIRS) verwendet, um zu untersuchen, wie domänenübergreifende und domänenspezifische Gehirnressourcen zur Komplexität beim dreistelligen Rechnen beitragen. Die Ergebnisse zeigten, dass die fronto-parietale Aktivierung mit der Komplexität des Rechnens zunahm, moderiert durch die Arbeitsgedächtniskapazität.

Zusammenfassend zeigt diese Dissertation, dass Rechnen mit Übertrag nicht immer gleichermaßen schwierig ist, sondern durch den Aufgabenkontext (Paradigma) sowie individuelle Einstellungen und Arbeitsgedächtnisressourcen beeinflusst wird. Das hat Auswirkungen darauf, wie Rechenfähigkeiten in der Bildungs- und Anwendungsbereichen erfasst und gefördert werden sollten.

Abbreviations

Abbreviation	**Definition**
ACC	anterior cingulate cortex
AG	angular gyrus
EEG	electroencephalography
fMRI	functional magnetic resonance imaging
fNIRS	functional near-infrared spectroscopy
IFG	inferior frontal gyrus
IPS	intraparietal sulcus
MFG	middle frontal gyrus
mDLPFC	middle portion of the dorsolateral prefrontal cortex
SMA	supplementary motor area
SMG	supramarginal gyrus

List of publications

Study 1:

Yao, X., Artemenko, C., He, Y., & Nuerk, H.-C. (2025). Arithmetic is not arithmetic: Paradigm matters for arithmetic effects. Cognition, 256, 106060. https://doi.org/10.1016/j.cognition.2024.106060.

Study 2:

Yao, X., Huber, J. F., Li, Z., Findik, Y., Nuerk, H.-C., & Artemenko, C. (2026). The dynamics of state math anxiety vary by paradigm and timing during arithmetic. *npj* Science of Learning. 11, 10. https://doi.org/10.1038/s41539-025-00398-z

Study 3:

Yao, X., Avcil, M., Meuer, P., Nuerk, H.-C., & Artemenko, C. Math self-concept decreases while math anxiety increases over the lifespan. Annals of the New York Academy of Sciences. (in revision)

Study 4:

Yao, X., Barth, B., & Artemenko, C. When arithmetic gets complex: fNIRS evidence for fronto-parietal activation in multi-digit arithmetic. *npj* Science of Learning (under review)

1. Introduction

1.1 General background: why arithmetic matters

In everyday life, there are many situations in which mental calculations are needed, like calculating the total price before buying something, verifying if a discount is right, or comparing prices of different items. Even if calculators or electronic devices are present, individuals tend to perform rapid mental calculations to arrive at a financial or mathematical judgment. Such simple calculations involve a number of basic mental operations that enable us to concentrate, hold information in mind, and coordinate numbers.

Therefore, arithmetic ability matters. Beyond its role in education and professional success, arithmetic ability predicts financial literacy, health-related decision-making, and even employment outcomes (Shomos & Forbes, 2014; Tushar & Sooraksa, 2023). However, arithmetic is not always an easy task for everyone. Although some simple problems can be solved by retrieving information from long-term memory, some people struggle when tasks become complex, particularly when multiple steps or intermediate transformations are required. In fact, complex arithmetic problems can cause anxiety and even math avoidance, which in turn can limit educational and career opportunities in a self-reinforcing cycle (Beilock & Maloney, 2015). Thus, the implications of arithmetic skills are far-reaching, yet what makes some arithmetic problems difficult and the mechanisms behind are still unclear.

Based on this background, this dissertation seeks to clarify what makes arithmetic problems “complex”, how complexity varies across contexts and individuals, and which processing demands increase as multi-digit complexity rises. Across four studies, this dissertation offer evidence on three main questions: (1) how task paradigms influence observed complexity effects and their associations with math attitudes; (2) how math attitudes vary on two timescales (moment-to-moment fluctuations within a task and age-related change across the lifespan) and how these variations relate to performance; and (3) how the fronto-parietal neural network underpins increasing carry/borrow demands in multi-digit arithmetic in a production paradigm.

1.2 Arithmetic complexity: definition and characteristics

Arithmetic becomes more complex when multi-digit numbers are involved. In multi-digit arithmetic, place-value integration involves carry/borrow operations that introduce multi-step transformations from units to decades. While arithmetic can also become complex when multiple operations are involved in a problem or when the representation of numbers goes beyond integers (fractions), the current dissertation is concerned with elementary addition and subtraction of integers because they provide a previse and controlled manipulation of complexity (via carry/borrow operations) and a well-established foundation for relating behavioral phenomena to underlying mechanisms.

1.2.1 Place-value system

Multi-digit arithmetic differs from single-digit arithmetic due to the place-value principle according to which the position of a digit determines its value in multi-digit numbers (e.g., the “3” in “37” represents 30, not 3). This creates additional processing demands in multi-digit tasks because successful performance requires extracting and integrating information across digit positions, and many arithmetic difficulties emerge when numbers and procedures extend beyond the single-digit range (e.g., Artemenko et al., 2024; Bahnmueller et al., 2018). Previously, researchers assumed that multi-digit numbers are represented holistically along a single mental number line, where each number specifies one particular location in a continuous magnitude representation (Dehaene et al., 1990; Restle, 1970). However, accumulating evidence from magnitude comparison tasks reveals that two-digit numbers are represented in a decomposed way, with decades and units processed separately, at least in addition to an overall holistic magnitude representation (Moeller et al., 2013; Nuerk et al., 2001; Nuerk & Willmes, 2005). This decomposed-but-interactive view is formalized in the general model framework proposed by Huber et al. (2016), which specifies how place-value components can be weighted and integrated in multi-digit processing. One prominent marker of decomposed processing is the unit-decade compatibility effect in two-digit number comparison: compatible number pairs where both decade and unit comparisons lead to the same decision (e.g., 42 vs. 57: 4 < 5 and 2 < 7) are processed faster and more accurately than incompatible pairs where the comparisons conflict (e.g., 37 vs. 62: 3 < 6 but 7 > 2), even when overall numerical distance between the numbers is matched (Nuerk et al., 2004). Importantly, the compatibility effect is one of several robust effects characteristic of multi-digit number processing (Nuerk et al., 2011). Although the unit digits are not necessary to determine the correct response in these stimuli (since the decade digits already determine which number is larger), they can nevertheless be processed and influence performance.

Supporting this decomposed framework, computational modelling studies comparing holistic, strictly decomposed, and hybrid models indicate that empirical data are often best simulated by componential models where separate magnitude representations exist for tens and units and flexible weighting of these components as a function of task demands (Huber et al., 2016; Moeller, Huber, et al., 2011). Moreover, eye-tracking evidence demonstrates that this decomposed processing occurs in parallel rather than sequentially, with both tens and unit digits influencing processing (Moeller et al., 2009). Importantly, the relative contribution of tens and units, and thus the size of compatibility effects, may be modulated by cognitive control settings rather than being purely automatic (Huber et al., 2014).

Given the decomposed processing of multi-digit numbers, Nuerk et al. (2015) proposed a three-level theoretical framework for understanding place-value processing: place identification, place-value activation, and place-value computation. Place identification refers to recognizing the position (e.g., units, tens) of individual digits within a multi-digit number. Place-value activation involves assessing the numerical magnitude of a digit based on its position (e.g., understanding the “3” in “37” as 30). Place-value computation includes operations that require calculations across different place-value columns, such as the carry or borrow procedures essential in multi-digit arithmetic. This framework offers a comprehensive perspective on the cognitive processes underlying multi-digit number processing.

Among the place-value computations (e.g., decomposing tens/units and integrating information across columns), carry and borrow operations represent particularly demanding processes that require coordinating information across multiple digit positions, making them important sources of complexity in multi-digit arithmetic.

1.2.2 Carry/borrow effects

Complexity in two- and multi-digit arithmetic increases when carry and borrow operations are required, that need place-value computation. Carry/borrow refers to transferring a decimal unit between adjacent positions. Carry operation in addition occurs when the sum of digits in a certain position exceeds 9, necessitating the transfer of one unit to the next higher position. For example, in 36 + 27, the unit sum (6 + 7 = 13) exceeds 9, requiring a carry of 1 to the decades position, such that the decades calculation becomes 3 + 2 + 1 = 6, yielding a result of 63. Borrow operation in subtraction is required when the minuend digit in a certain position is smaller than the corresponding subtrahend digit, necessitating borrowing a decade from the next higher position. For example, in 63 – 27, 3 < 7 in the unit position, so that one decade must be borrowed from the decades' position, transforming the unit's calculation to 13 – 7 = 6 and the decade's calculation to 6 – 1 – 2 = 3, yielding a result of 36.

The carry and borrow operations increase task demands, with problems requiring carry or borrow operations consistently showing longer reaction times and higher error rates compared to problems without such operations, which has been known as the carry effect in addition and the borrow effect in subtraction (Artemenko, 2018; Imbo et al., 2007; Moeller, Klein, et al., 2011b). These effects reflect the additional cognitive load, e.g., for maintaining intermediate results in working memory (Imbo et al., 2007b), coordinating operations across place-value positions (Lambert & Moeller, 2019), and executing multi-step procedures rather than single retrievals (Ding et al., 2019).

While two-digit problems typically involve at most one carry or borrow, three-digit problems can necessitate multiple sequential transfers across columns. Each additional transfer operation increases the demands on working memory (Imbo et al., 2007a). For example, 152 + 234 (no carry), 158 + 234 (one carry from units to tens), and 158 + 567 (two carries from units to tens and from tens to hundreds). Each additional operation increases the number of elements that must be simultaneously maintained and updated in working memory (Hawes et al., 2019), resulting in higher difficulty. Similarly, in three-digit subtraction like 534 – 278, borrow may cascade from the hundreds to the decades and then to the unit's position when the subtrahend digits are larger than the corresponding minuend digits. Behavioral studies demonstrate that the presence of a carry operation in multi-digit addition directly predicts longer reaction times and higher error rates (Klein, Moeller, et al., 2010). These cascading operations require not only maintaining multiple intermediate results in working memory but also tracking the sequence of transformations across positions, substantially increasing the procedural complexity compared to problems with single or no carry/borrow operations (Imbo et al., 2007b).

Carry and borrow operations introduce complexity in both categorical and continuous ways (Klein, Willmes, et al., 2010). Categorical complexity refers to whether any carry/borrow is required at all, whereas continuous aspects capture graded increases in difficulty, for instance with larger unit sums (Klein, Willmes, et al., 2010; Lambert & Moeller, 2019). Eye-tracking and behavioral work shows that carry is detected early during problem encoding and that each additional transfer step entails further updating and coordination in working memory (Ding et al., 2019; Moeller, Klein, et al., 2011b). While categorical complexity relates to procedural and problem-solving processes associated with language areas and basal ganglia (Delazer et al., 2004; Menon et al., 2000), continuous complexity, which includes magnitude processing, mainly engages the intraparietal regions responsible for magnitude representation and place-value integration (Sokolowski et al., 2023), reflecting increased demands on numerical manipulation and working memory updating (Viesel-Nordmeyer & Prado, 2023). The behavioral expression of carry/borrow demands may also vary with the strategies elicited by a given task. Eye-tracking evidence suggests that adults may take distractor information into account more than younger children during carry addition (Moeller, Klein, et al., 2011a). Consistent with developmental differences in strategy use, Caviola et al. (2018) reported that older children more often relied on efficient memory-based and decomposition strategies, whereas younger children more frequently used counting and right-to-left procedural algorithms, with strategy selection varying as a function of age, problem complexity, and presentation format.

In sum, these findings indicate that carry and borrow steps are a main source of arithmetic complexity because they require coordination between working memory control and numerical magnitude processing. Yet most of this work has relied on relatively single-/two-digit problems in decision paradigms, leaving open how these sources of complexity manifest in more naturalistic multi-digit production tasks and in the brain.

1.2.3 Addition versus subtraction

While subtraction and addition share core place-value computations, they are not cognitively interchangeable (Artemenko & Nuerk, 2025). Subtraction can be viewed as the inverse operation of addition, which may require additional inversion-specific processing beyond the computations shared with addition. The inversion account suggests that subtraction performance reflects a combination of shared place-value computation and operation-specific inversion demands. This explains why subtraction, in particular borrow operation, often imposes greater processing demands than addition and carry operation.

Consistent with this framework, subtraction was found to show slower reaction time and higher error rate than addition (Artemenko, 2018; Campbell, 2008). Educational studies confirm that students find subtraction harder, with errors persisting even in older learners (Narciss & Huth, 2006). At the neural level, neuroimaging studies have reported partly different activation patterns for subtraction versus addition. Specifically, subtraction may engage more bilateral activation in fronto-parietal regions, whereas addition sometimes shows more left-lateralized patterns, although results vary across studies and appear sensitive to problem complexity and strategy use (for a meta-analysis see Istomina & Arsalidou, 2024; Rosenberg-Lee et al., 2011). Yang et al. (2017) employed fMRI and demonstrated that subtraction elicited stronger activation in the left inferior frontal gyrus (IFG), middle portion of the dorsolateral prefrontal cortex (mDLPFC), and supplementary motor area (SMA) compared to addition. The connectivity analysis revealed that subtraction engaged bilateral intraparietal sulcus (IPS) circuits, with a particular emphasis on the right IPS for magnitude processing, whereas addition tended to rely on left-hemisphere pathways for retrieval. Likewise, another research found that compared with addition, complex subtraction activated additional right-hemisphere areas (precentral cortex and thalamus) and left inferior parietal lobule, whereas complex addition more strongly activated medial frontal cortex (Yi-Rong et al., 2011). Artemenko et al. (2018a) also found operation-specific neural patterns that vary with complexity and developmental stage, suggesting that the cognitive mechanisms underlying addition and subtraction share similarities but also with distinctions. In sum, these findings align with the inversion account in that operation differences are conditional rather than fixed. Subtraction may appear more demanding than addition when inversion-related transformations and cross-column coordination are required. However, when faced with simple problems without carry or borrow, performance and neural activation may converge because both operations showed shared arithmetic network whose engagement increasing with processing demands (Artemenko & Nuerk, 2025).

Importantly, much of the evidence informing the inversion account is derived from two-digit problems that involve at most one carry/borrow step. It therefore remains an open question whether the finding of larger cost for borrow than carry extends to more complex multi-digit arithmetic, such as three-digit addition and subtraction, where more than one carry/borrow operations maybe needed and may increase demands on cross column coordination.

1.3 Task contexts: do different paradigms matter?

Besides the arithmetic task itself, the task context in which the problem is presented and solved may also influence arithmetic performance. For example, studies showed that arithmetic problems presented in a vertical format were solved faster than in a horizontal format. Moreover, problems with double digit appeared as the first addend (e.g., 52 + 3) were solved more quickly than the single digit appeared as the first added (e.g., 3 + 52) when problems were presented horizontally, indicating that these different presentation format can influence performance by engaging different working memory components (Trbovich & LeFevre, 2003).

Beyond the directional presentation format, experimental paradigms also work as task contexts. Experimental paradigm refers to the specific task format and response requirements used to study arithmetic. Although complex arithmetic tasks involve fronto-parietal network, the behavioral performance of arithmetic effects (e.g., carry/borrow effect) may depend on the paradigm used. This matters because laboratory paradigms often differ from everyday arithmetic.

Neuroimaging studies on the neural correlates of arithmetic usually use verification or forced-choice paradigms due to methodological restrictions by functional magnetic resonance imaging (MRI) or electroencephalography (EEG) (De Smedt et al., 2013; Grabner et al., 2009; Hinault & Lemaire, 2016; Tschentscher & Hauk, 2014). Because of the easier experimental control, even behavioral studies mostly use such paradigms (J. I. Campbell & Fugelsang, 2001; Thevenot et al., 2020; N. J. Zbrodoff, 1999). In verification or forced-choice paradigms, one or multiple solution probes are presented together with the arithmetic problem and the subjects are required to indicate whether a given solution is true or false (e.g., 26 + 53 = 69; verification paradigm) or to choose the correct solution probe (e.g., 26 + 53 = 69 or 79; forced-choice paradigm). Although verification or forced-choice tasks are widely used for their convenience, evidence suggests they engage different cognitive processes from production tasks. Individuals may assess an equation as a whole and often rely on familiarity or reject distractors, rather than compute the correct answer explicitly (Ashcraft & Stazyk, 1981; N. J. Zbrodoff & Logan, 1986). Verification has been argued to rely partly on evaluating the overall “resonance” or match the equation with memory representations, rather than strictly following a “production plus comparison” sequence (Zbrodoff & Logan, 1990). Eye-tracking evidence from forced-choice paradigm suggests that distractors are actively processed, demonstrating that performance in such tasks extends beyond pure computation of the correct answer (Moeller, Klein, et al., 2011a). Specifically, forced-choice paradigms may encourage distractor-driven shortcuts. When distractors are constructed as small deviations from the correct result (e.g., from ±1 to ±9), participants may sometimes respond based on partial consistency checks (such as the unit digit), rather than fully computing the problem and integrating carry/borrow operations across place values.

By contrast, arithmetic in educational and everyday contexts is often performed in production paradigms, where an answer must be generated without external probes. Although rarely used, arithmetic research with production paradigms is possible (Artemenko, Soltanlou, Ehlis, et al., 2018; Dewi et al., 2021). In a production paradigm, subjects are required to solve the arithmetic problem mentally, and to respond by writing, typing or saying out the solution, thereby requiring step by step computation without shortcuts. Since the different paradigms may involve different cognitive processes, accordingly, an open question, is to what extent paradigm choice influences behavioral performance and the presence of arithmetic complexity effects (e.g., the carry or borrow effect)? This matters because if performance in different paradigms largely depends on the paradigm used, then conclusions from one paradigm couldn't be generalized to other paradigms interchangeably.

1.4 Individual differences in arithmetic complexity

Differences in paradigms may involve different cognitive processes and evaluative context in which arithmetic is performed, which in turn may influence individuals' affective responses during the task. This raises the question of whether paradigm-dependent contexts are accompanied by systematic differences in math attitudes. Within the tripartite models of math attitudes (Wen & Dubé, 2022), math anxiety and math self-concept are components of mathematics attitudes, representing affect and self-evaluation, respectively. Accordingly, the present dissertation focuses on math anxiety and math self-concept and examines how state math anxiety fluctuates within tasks (micro-timescale dynamics) and how math attitudes vary across the lifespan (macro-timescale change), providing a temporal perspective on when and for whom arithmetic complexity effects are most evident.

1.4.1 Micro-timescale: within-task dynamics of state math anxiety

Math anxiety has been widely recognized as a barrier to math achievement. It is defined as "a feeling of tension and anxiety that interferes with the manipulation of numbers and the solving of mathematical problems in ordinary life and academic situations" (Richardson & Suinn, 1972). It manifests as fear or apprehension when one is faced with numerical tasks, ranging from basic arithmetic to more advanced problem solving. Math anxiety is common among students, the majority report moderate to high levels of anxiety specifically related to math, sometimes even in early school years (Dowker et al., 2016).

Math anxiety can be understood within the broader state–trait anxiety framework (Spielberger, 2013). Trait math anxiety refers to a long-term, generalized predisposition to feel anxious about math, typically measured by self-report of one's usual anxiety in math contexts. In contrast, state math anxiety is a temporary, situation-specific feeling of anxiety that arises during a particular math task or moment (Roos et al., 2015). For example, a student might generally consider themselves comfortable with math but still experience a spike of acute anxiety when faced with a challenging problem under time pressure. High state math anxiety consumes working memory resources, impairing performance, while trait math anxiety, which is often associated with more persistent worry and avoidance, has shown varied associations with academic achievement (Cipora et al., 2022; Orbach et al., 2019).

Math anxiety is more than just a general dislike of math, it has measurable detrimental effects on performance (Barroso et al., 2021; Brunner et al., 2023). There is significant evidence that math anxiety interferes with the execution of mathematical tasks, especially those that rely on working memory (Ashcraft & Kirk, 2001). Based on the processing efficiency theory, the working memory capacity is occupied by anxious thoughts and worries, resulting in reduced capacity for actual calculation or problem solving (Eysenck & Calvo, 1992). Empirical evidence also supports this process, such that people who experience higher levels of math anxiety tend to perform significantly worse on complex arithmetic problems that require carry or borrow operations and the maintenance of intermediate results or strategies, but not on simple problems that do not require carry or borrow operations (Huber & Artemenko, 2021). Moreover, empirical findings have revealed that state and trait math anxiety can present different patterns of prediction, where state math anxiety was found to be related to easier math problems, whereas trait math anxiety was found to be related to more difficult math problems (Demedts et al., 2022). This further emphasizes the importance of distinguishing between "in-the-moment" anxiety and trait anxiety.

However, most research on math anxiety and performance measures state math anxiety only once before or after a task, and how such dynamics relate to different task phases remains largely unexplored (Conlon et al., 2021), leaving open question of how state math anxiety changes across different phases of arithmetic tasks, e.g., when the state math anxiety peaks and when it decreases.

1.4.2 Macro-timescale: lifespan differences in math attitudes and their performance links

At a broader timescale, evidence from personality suggests that mean levels of neuroticism tend to decline across adulthood (Specht et al., 2011; Wortman et al., 2012). Moreover, aging was found to be associated with reduction in susceptibility of anxiety (Jorm, 2000). These age-related changes raise a question for math domain: whether and how math attitudes change with aging. If math attitudes change across adulthood, then associations between math attitudes and arithmetic performance that are primarily derived from students or younger adult samples may not generalize to older adults.

Math self-concept is defined as an individual's self-perception of his/her own competence and confidence in math, which plays a crucial role in influencing motivation and engagement in math-related activities (Jacobs et al., 2002; Wigfield & Eccles, 2000). It has been found that a positive math self-concept is associated with greater achievement in math (e.g., Wang, 2023). Students who see themselves as "good at math" are more likely to tackle challenging problems, use effective strategies, and persist when faced with difficulties, which helps to improve learning and performance. On the other hand, students with a low math self-concept may question their ability, resulting in less effort or giving up more quickly, regardless of their actual ability (Passiatore et al., 2024).

Math self-concept and math anxiety are related but distinct: a student may experience math anxiety even if they recognize that they are generally good at math, or vice versa, a student may believe they are "not a math person" (low self-concept) but not experience math anxiety (Marsh & Craven, 2006). Longitudinal studies have shown that these two constructs have a reciprocal relationship, in that lower math self-concept predicts higher math anxiety, and higher math anxiety decreases future self-concept (Ahmed et al., 2012). This implies that interventions designed to increase learners' confidence and sense of agency in math may also decrease math anxiety and its negative impacts. Importantly, math self-concept was shown to be a more significant driver of crucial academic and career decisions than math anxiety alone, suggesting its ecological validity in understanding individuals' long-term avoidance and engagement with mathematics (Lunardon, et al., 2025).

These math attitudes may be especially relevant when arithmetic becomes more complex. Multi-digit problems involving carry and borrow operations impose high demands on working memory and cognitive control, and prior work suggests that math anxiety can impair performance particularly when arithmetic becomes more complex (Huber & Artemenko, 2021). This raises a related question: beyond math anxiety, do individual differences in math self-concept likewise modulate how individuals are affected by arithmetic complexity? According to expectancy-value theory, individuals' competence-related beliefs (including self-concept) can influence their achievement-related choices, persistence and performance. Correspondingly, math self-concept was found to be associated with students' willingness to engage with and persist on challenging math tasks (Jacobs et al., 2002; Marsh & Craven, 2006; Passiatore et al., 2024; Wigfield & Eccles, 2000). Moreover, previous research indicated that negative association between math anxiety and math performance was explained partly by math self-concept together with working memory (Justicia-Galiano et al., 2017), suggesting that math self-concept may also influence performance especially on complex arithmetic.

A lifespan approach proposes that basic arithmetic skill improves markedly through childhood into early adulthood and stabilizes in adulthood, even as domain-general cognitive abilities (such as working memory and processing speed) decrease with age (Avcil & Artemenko, 2025). At the same time, math attitudes emerge in childhood, with math anxiety and math self-concept emerging in early schooling (Gunderson et al., 2012), and math anxiety appearing as early as first grade and relating negatively to math achievement (Ramirez et al., 2013). Social learning environment shapes these. For example, math anxiety in parents can be a strong predictor of math achievement and math anxiety in their children, especially when parents who are math anxious often help with math homework (Maloney et al., 2015). Teachers' own anxiety is important: primary teachers tend to be more math anxious than other adults, which can inadvertently transmit math anxiety to students (Artemenko et al., 2021; Cipora et al., 2024).

Despite the rich evidence during school, less is known about how math attitudes evolve across adulthood, especially middle-aged or older adulthood. It is important to understand lifespan macro-changes because attitudes at any age can affect continued math use, learning, and performance.

1.5 Neural substrates of arithmetic complexity

Previous sections have illustrated how arithmetic complexity arises from place-value integration and carry/borrow demands as well as influenced by task contexts and attitudinal factors. Yet behavioral data alone cannot fully explain how these demands are processed in the brain. To understand the mechanisms supporting multi-digit arithmetic, it is necessary to identify the underlying neural mechanisms. This section reviews theoretical models and neural evidence for the fronto-parietal network supporting arithmetic processing.

1.5.1 Theoretical frameworks

The Triple Code Model of number processing (Dehaene & Cohen, 1997) proposes that numbers are encoded in three different formats: a visual code (Arabic digit strings), a verbal code (spoken/written number words), and an analog magnitude code (a mental number line). These are supported by partially overlapping but distinct areas of the brain. For example, visual digit recognition involves occipito-temporal areas (visual number form), verbal codes involve language areas, and magnitude processing engages the bilateral intraparietal sulci (IPS). The Triple Code Model has been successful in explaining many results (e.g., double dissociations between fact retrieval and magnitude comparison) and has led to further developments of the proposed parietal circuits for numerical magnitude processing (Dehaene et al., 2003).

However, neuroimaging and meta-analytic studies also suggest that there is a region of overlap between these codes in a common fronto-parietal network, where number comparison, calculation, and symbol processing often engage regions of overlap (Arsalidou & Taylor, 2011; Nieder & Dehaene, 2009). This leads to the consideration of an effect-based approach that investigates the impact of particular operations in multi-digit arithmetic (such as carry and borrow) on the recruitment of regions, independent of task or format differences. In practice, the Triple Code Model does not explicitly specify how these codes work together during complex multi-digit arithmetic or domain-general resources contribute.

Based on Triple Code Model, Klein et al. (2016) proposed an extension comprising two coupled fronto-parietal networks: a dorsal magnitude network and a ventral arithmetic-fact network, which are connected through fronto-parietal fiber pathways. This Two-network Framework predicts that tasks requiring extensive quantity manipulation and procedures would activate the magnitude network more, while those relying on rote fact retrieval would activate the fact network (Klein & Knops, 2023). Importantly, the Two-network Framework highlights that these two networks work in a context-dependent manner together: even simple problems may recruit some magnitude processing, and difficult problems increasingly turn to magnitude processing over fact retrieval. In other words, as problems become more difficult, the contribution of the magnitude network would increase, and the contribution of fact network would decrease. The Two-network Framework explains connectivity issues (e.g., the role of the hippocampal in learning facts) and developmental issues (how these two networks develop over time), but it also raises questions about how the sequential place-value processing is carried out in these two networks.

In sum, the Triple Code Model and the Two-network Framework both emphasize arithmetic in the fronto-parietal networks, but with a different focus. Taken together, these models provide the theoretical foundation for understanding the processing of arithmetic complexity: These models specifically integrate domain-specific representations for numbers (e.g., place value, magnitude) with domain-general control and working memory processes in the fronto-parietal networks. Based on this integrated framework, the following section focuses on empirical evidence for domain-general and domain-specific contributions to multi-digit arithmetic and uses this evidence to motivate a further investigation of how complex, multi-step carry or borrow operations are implemented processed in the fronto-parietal networks.

1.5.2 Domain-general and domain-specific processes in the fronto-parietal network

Arithmetic performance is a reflection of both contribution of domain-general resources and domain-specific numerical representations, and therefore, to understand arithmetic, one must take into account both domain (Knops et al., 2017). Neuroimaging studies concentrate on a distributed fronto-parietal network that processes domain-general tasks mostly in frontal brain regions and domain-specific numerical demands mostly in parietal areas of the brain (Menon, 2016; Sokolowski et al., 2023). The lateral prefrontal cortex (specifically, the inferior and middle frontal gyri, IFG and MFG) and medial frontal areas (anterior cingulate, ACC; supplementary motor area, SMA) are involved in working memory, sequential planning, top-down control, conflict monitoring, and task initiation for multi-step calculations (Menon, 2016). In the parietal cortex, the intraparietal sulcus (IPS) and superior parietal lobule represent numerical magnitude and place-value relationships (Nieder & Dehaene, 2009). The inferior parietal regions, including the angular gyrus (AG) and adjacent supramarginal gyrus (SMG), are particularly active during the retrieval of well-practiced arithmetic facts (Polspoel et al., 2017). The fronto-parietal network is known to support both the manipulation of quantities and the retrieval of numeric knowledge that are both involved in arithmetic.

Domain-general cognitive resources (e.g., working memory, attention, and cognitive control) are recruited via frontal and cingulate circuits to manage intermediate results and sequence steps (Botvinick et al., 2004; Miller & Cohen, 2001). From an effect-based perspective, carry and borrow are useful for investigating arithmetic mechanisms because they increase complexity by adding intermediate steps and place-value updating demands. For example, carry in addition or borrow in subtraction impose a heavy working memory demand (Imbo et al., 2007b; Imbo & LeFevre, 2010) and selectively increase activation in lateral prefrontal regions, especially left IFG and bilateral MFG (Artemenko, Soltanlou, Dresler, et al., 2018). Converging neuroimaging studies report stronger activation in lateral prefrontal and cingulate regions for problems that require carry or borrow operations than for no-carry/borrow problems, including ACC and SMA, consistent with increased cognitive control demands during these steps (Artemenko, Soltanlou, Dresler, et al., 2018; Klein, Willmes, et al., 2010; Kong et al., 2005).

Domain-specific numerical processing involves the representation numerical magnitude and place-value, subserved mainly by parietal regions. Arithmetic problems with larger operands (the "problem size" effect) has been shown to reliably enhance IPS activity, which is thought to be due to the increased demands of quantity processing (Prado et al., 2013). More specifically, whereas the categorical aspects of the carry effect are associated with frontal and subcortical regions, the continuous aspects (e.g., unit-sum) are reflected in parietal activity, namely by the increased activation of IPS and superior parietal areas (Klein, Willmes, et al., 2010). Nevertheless, the contribution of increased parietal activity, particularly in the IPS, to the performance of the categorical carry/borrow operations is not universally observed; for instance, Artemenko et al. (2018) found that when problem size is controlled, the effect is mainly linked to frontal areas, which are thought to be involved in domain-general cognitive processing.

In sum, these studies indicate that carry and borrow operations require domain-general control circuits to work together with domain-specific place-value representations. However, most neuroimaging studies on arithmetic have focused on single-digit or, at most, two-digit problems, usually involving verification tasks. Two-digit can show that arithmetic engages fronto-parietal regions, but they can’t fully reveal the detailed computation processes when more place-value integration steps are involved. While in three-digit arithmetic, more than one carry/borrow operation can be required, making the computation process more complex. However, we know relatively little about how the brain supports multi-digit arithmetic, especially when solving three-digit problems that involve multiple carry/borrow operations and require place-value integration in ecological production settings.

2. Research aims and scope

Building on the preceding sections, the current dissertation investigates arithmetic complexity in multi-digit calculation by concentrating on the role of carry and borrow operations as a major source of place-value integration complexity. In the four preregistered studies, arithmetic complexity is defined by the presence (two-digit tasks in Study 1, 2, and 3) and, in three-digit arithmetic, the number of carry/borrow steps, while controlling for problem size. This work tackles arithmetic complexity from three different perspectives: task paradigm, math attitudes, and fronto-parietal neural mechanisms. The dissertation consists of four empirical studies:

Study 1: Experimental Paradigms.

Arithmetic research in the laboratory uses a variety of paradigms that are often treated as interchangeable, but there could be some differences that affect cognitive load and, therefore, arithmetic effects (such as carry and borrow effects). For example, generating an answer, which is common in classroom, usually requires more working memory resources and could cause higher levels of anxiety compared to simply recognizing the correctness of a given solution. This study systematically compares performance on two-digit addition and subtraction across widely used paradigms to test whether paradigms influence performance and the observed arithmetic effects. This is important as a guidance for future studies when choosing paradigms in arithmetic research.

Study 2: Trait and State Math Anxiety. Trait math anxiety reflects individual's general tendency to experience anxiety in math-related contexts, while state math anxiety refers to the momentary emotional reactions triggered during specific tasks. However, it remains unclear how state math anxiety fluctuates during arithmetic tasks and whether it depend on the paradigms. This study investigated the interaction between the two types of math anxieties and the paradigms used in two-digit addition and subtraction tasks. Specifically, it explored whether production paradigms, which require participants to generate answers, are associated with higher levels of state math anxiety than decision paradigms. Additionally, this study measured the variation in state math anxiety across the phases of the tasks, which captures the temporal dynamics of anxiety before, during, and after arithmetic performance. This study combines self-report questionnaires and task performance measures to provide a picture of how trait dispositions and momentary emotional states interact to influence performance.

Study 3: Lifespan Development. Given the importance of arithmetic across the lifespan and the fact that math attitudes may impact arithmetic performance, it is necessary to comprehend the impact of these variables across the lifespan. Math attitudes change with age are plausible given the fact that domain-general resources such as processing speed and working memory have been found to change across the lifespan. However, most of the previous studies have focused on children or young adults, and thus there is a gap in the current literature regarding the impact of math anxiety, math self-concept, and arithmetic performance across the lifespan. Using cross-sectional samples ranging from childhood to older adulthood, this study takes a lifespan approach to comprehend the impact of math attitudes and their association with the complexity effect (carry/borrow effect) in two-digit arithmetic.

Study 4: Neural Correlates. Whether increasing complexity engages domain-general processes, such as working memory in frontal regions, or also recruits domain-specific processes, such as magnitude and place-value processing in parietal regions, remains unknown. To answer this question and relate behavioral observations to neural mechanisms, the current study investigates the brain activity associated with multi-digit arithmetic complexity. Extending the two-digit arithmetic complexity investigated in Studies 1-3, this study investigates cortical activation during three-digit addition and subtraction problems with no, one, or two carry or borrow operations, and whether increased difficulty is more related to the engagement of domain-general frontal regions or also involves the engagement of domain-specific parietal areas. To allow ecologically valid measurements, this study uses functional near-infrared spectroscopy (fNIRS), a movement-tolerant neuroimaging technique suitable for naturalistic production paradigms. This contrasts with traditional neuroimaging studies, which have largely focused on single-digit arithmetic and verification or forced-choice paradigms. This study aims to provide neural evidence complementing behavioral data and contribute to the understanding of how domain-general and domain-specific brain network work to support arithmetic processing.

Together, these four studies provide an integrated investigation of arithmetic complexity across paradigm, math attitudes, and fronto-parietal neural mechanisms (See Figure 1 and Table 1 for an overview of four studies).

Figure 1. Graphical overview of research questions in each study

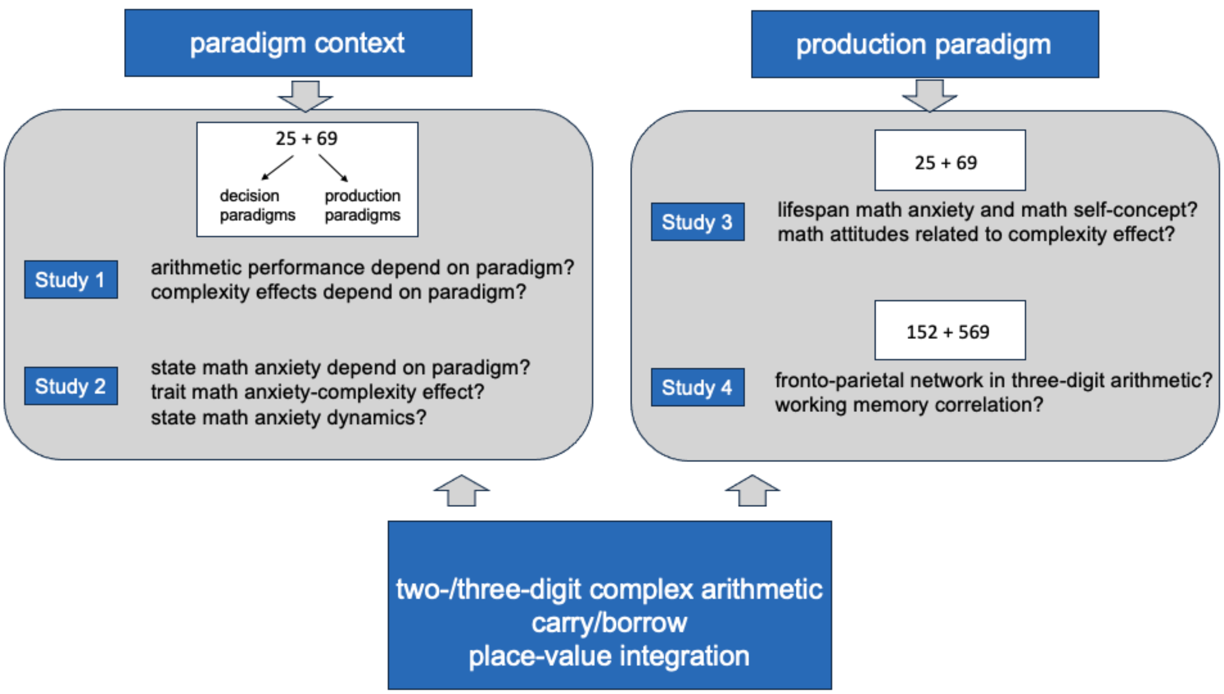


Table 1. Overview about details across four studies

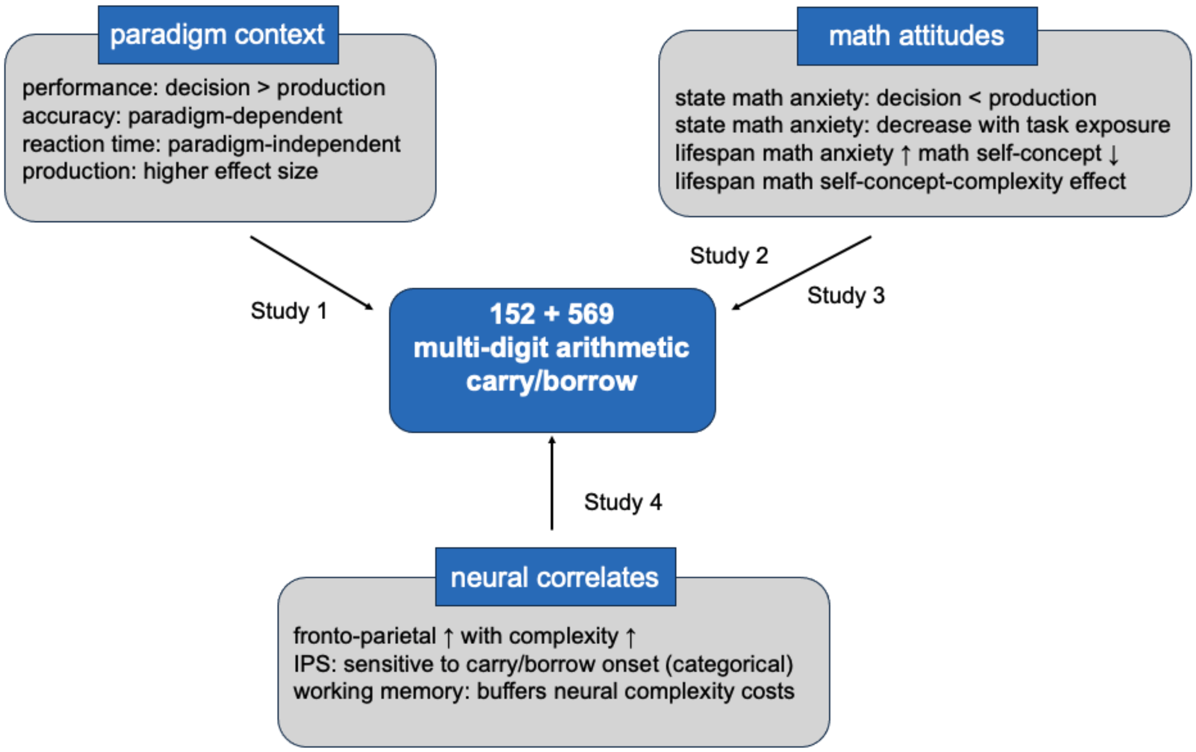
	Study 1	**Study 2**	**Study 3**	**Study 4**
Tasks	two-digit arithmetic			three-digit arithmetic
Sample	*N* = 65 adults		children (*N* = 53); younger adolescents (*N* = 50); older adolescents (*N* = 50); younger adults (*N* = 53); middle-aged adults (*N* = 50); older adults (*N* = 50)	*N* = 55 adults
Method	• behavioral experiment	• behavioral experiment • questionnaire	• behavioral experiment • questionnaire	• behavioral experiment • fNIRS
Variables	• complexity (with vs. without carry/borrow) • operation (addition vs. subtraction) • paradigm (6 different paradigms)	• complexity (with vs. without carry/borrow) • operation (addition vs. subtraction) • paradigm (decision vs. production)	• complexity (with vs. without carry/borrow) • operation (addition vs. subtraction)	• complexity (0,1,2 carry/borrow operations) • operation (addition vs. subtraction)

Measures	• performance (reaction time and accuracy)	• performance (reaction time and accuracy) • trait math anxiety • state math anxiety	• performance (reaction time and accuracy) • trait math anxiety • math self-concept	• performance (reaction time and accuracy) • working memory
Open Science	preregistered; data, materials, analysis are openly shared			

3. General discussion

This dissertation explores arithmetic complexity in multi-digit calculations in four studies on behavioral performance, task paradigms, individual differences in math attitudes, and neural mechanisms. The overall results suggest that complexity effects are mainly caused by carry and borrow operations, and that complexity effects are consistently observed in reaction time instead of accuracy, which were more dependent on the task context and individual differences. In Study 1, accuracy was generally higher in decision paradigms, especially verification, which might mitigate complexity effects in accuracy. As expected, Study 2 found that the impact of math anxiety on performance is mainly seen in reduced processing efficiency (slower reaction time) but not in impaired accuracy, and that production paradigms are associated with higher state math anxiety, which decreases with task exposure. Extending to the perspective of lifespan, Study 3 found that math self-concept declines and math anxiety increases with age. Importantly, math self-concept was negatively related to complexity effect in error rate. Neuroimaging Study 4 showed that fronto-parietal network activity increased with complexity, and that this neural complexity effect was differentially buffered by individual working memory capacity. Discussions about these results are as follows (see Figure 2 for the overview of findings).

Figure 2. Graphical overview of findings



Notes. The symbol "↑" means increasing, and "↓" means decreasing.

3.1 Arithmetic complexity： carry and borrow

Across the four studies, complexity effects were observed in tasks requiring carry/borrow operations in both two- and three-digit calculation. Compared to simple arithmetic without carry/borrow operations, the complex arithmetic involving carry/borrow operations require place-value integration and introduces extra cognitive demands, e.g., higher working memory involvement (for maintaining and updating carried or borrowed values) and more fronto-parietal activation (e.g., Study 4 & Imbo et al., 2007a). In our research, the problem size was controlled in both two- and three-digit arithmetic tasks, thereby reducing the extent to which numerical magnitude alone can explain for the observed complexity effect. Instead, the worse performance in complex tasks reflects both the domain-general (working memory) and domain-specific place-value processing demands (Artemenko & Nuerk, 2025).

For three-digit arithmetic in Study 4, varying the number of carry/borrow operations makes it possible to separate categorical from continuous complexity. The transition from no carry/borrow to a single carry/borrow captures a categorical component, because participants must detect whether a carry/borrow operation is required and initiate the corresponding place-value procedure (Klein, Moeller, et al., 2010). Previous research showed that this kind of categorical decision was driven by domain-general central executive function, which can work for managing and coordinating the yes or no determination of carry or borrow (Artemenko, Soltanlou, Ehlis, et al., 2018; Imbo et al., 2007b). While Klein et al. (2010) identified a continuous characteristic of the carry effect in two-digit addition, demonstrating that reaction times and intraparietal activation increased continuously with the magnitude of the unit sum, our research extended the continuous patterns into three-digit arithmetic including carry/borrow. When tasks involve two successive carry/borrow operations, this manipulation allows us to test whether complexity continues increasing with the number of carry/borrow steps, beyond the initial categorical decision that a carry/borrow is required, by increasing procedural and control demands (Imbo et al., 2007a). Within the framework, our fNIRS data showed that the domain-specific right intraparietal sulcus (IPS) activation increased from no carry/borrow to one carry/borrow but did not rise further from one carry/borrow to two carries/borrows, indicating a categorical sensitivity of the right IPS to the detection of carry/borrow operation. However, the lack of an increase should be interpreted with caution, considering the spatial specificity and sensitivity constraints of fNIRS.

In conclusion, the control of problem size points to the carry/borrow operation as a source of complexity in multi-digit arithmetic, and this will be discussed in the next section.

3.2 Operation effect: subtraction is inversion of addition

In the present dissertation, both study 1 (two-digit arithmetic) and study 4 (three-digit arithmetic) revealed operation effect, with subtraction yielded longer reaction time than addition, and in Study 4 it also resulted in lower accuracy. This patterns supports the inversion account (Artemenko & Nuerk, 2025), which proposes that subtraction is the inverse computation of addition and this inversion brings extra processing steps and cognitive demands. Such demands increase reliance on working memory for managing intermediate calculations (Imbo et al., 2005), and may encourage strategy shift of using indirect addition (Peters et al., 2010), or the less automatized backward operation on the mental number line (GrÉGoire & Van Der Linden, 1997). Taken together, the operation effect observed here can be interpreted as an inversion-related procedural burden: subtraction relies on the same place-value framework as addition, but borrow operation introduces extra inversion-linked steps and place-value integration demands, which increases reaction time and error rates, as the majority of errors in multi-digit arithmetic are due to the carry/borrow operations and place-value integrations (Artemenko & Nuerk, 2025).

At the neural level, study 4 did not show a main effect of operation, nor an operation and complexity interaction, within the recorded fronto-parietal fNIRS channels, despite robust complexity effects, this result also aligns with another research on word problems using verbal production paradigm (Daroczy et al., 2026). This could be viewed as a boundary condition for how operation effects are expressed in this paradigm. In our cases, with arithmetic task magnitude matched in the production paradigm, addition and subtraction may rely on a shared fronto-parietal network that supports the place-value computation. While the subtraction cost is expressed in the inversion process of intermediate states rather than a stable mean amplitude difference between operations within the measured channels. This interpretation corresponds to Study 1, where the operation effect in reaction time was replicated across all six paradigms while the accuracy operation effect appears mainly in verbal production paradigms and was attenuated in decision paradigms. A plausible reason is that given answer options in decision paradigms allows for strategies or shortcuts which were not assessed in our current research. Taken together, the behavioral and neural pattern suggests that operation differences in multi-digit arithmetic can be robust on behavioral performance level while remaining difficult to detect in neural level in our fNIRS study, especially when the task constrains strategies and emphasizes computation and produce answers in our adopted production paradigms.

Moreover, previous evidence has shown that children exhibit low levels of flexibility in the strategies they actually employ during multi-digit subtraction and low levels of adaptability in selecting optimal strategies (Hickendorff, 2022). Interestingly, beyond arithmetic, research on problem solving also revealed the operation effects: people tend to systematically use additive transformation and overlook subtractive transformations unless the task environment cues subtraction (Adams et al., 2021). These findings are consistent with the inversion account (Artemenko & Nuerk, 2025) in suggesting that subtractive operations impose additional demands both on arithmetic and on problem-solving perspective. Besides, the problem-solving study also reminds us of the importance of external context, which we will discuss in the next section.

3.3 Paradigm is a context variable

An important finding in the dissertation is that paradigm is not a neutral methodology, whereas it is a context variable and can systematically influence individuals' performance and affective response. Study 1 and Study 2 offer complementary evidence from cognitive and affective perspectives: Evidence in Study 1 demonstrated higher accuracy in decision paradigms while production paradigms are more prone to errors, with complexity effect and operation effect in reaction time showed consistent patterns across six paradigms. This suggests that accuracy is paradigm dependent. In parallel, Study 2 found that state math anxiety was greater in production paradigms than decision paradigms, and this paradigm difference was greater for those with higher trait math anxiety. Decision paradigms provide options and allow strategy use such as plausibility checking or rejecting wrong options instead of calculating step by step. However, production paradigms measure pure calculation tasks with no external cues or options. In this case, producing an answer may increase individuals' uncertainty about their answers as well as worries about making mistakes, these all contribute to the increased state math anxiety in production paradigms.

The paradigm-dependent performance and paradigm-dependent state math anxiety together provide implications for answering "which is a better paradigm": When research concerns accuracy or error patterns of the arithmetic as well as state affective responses, paradigms should be chosen with caution since they can influence the results as well as related conclusions. While if research concerns more on speed of arithmetic, then different paradigms may produce comparable conclusions. For example, in the case of analyzing arithmetic errors in financial contexts in Parkinson's disease, the production paradigms with open responses made it possible to identify and categorize specific error types, especially those involving place-value integration (Loenneker et al., 2021). Besides, the Trier Social Stress Test involving arithmetic tasks under strict time limits are often used as cognitive stressors in psychological research (Kirschbaum et al., 1993; MacCormack et al., 2024). In such cases, presenting arithmetic tasks in production paradigms is likely to be more effective in eliciting stronger stress than decision paradigms.

Notably, context is not limited to paradigms in arithmetic tasks. Whereas paradigms mainly differ in the response formats, language in word problems change what must be understood and represented before calculation can begin. Previous research have shown that text processing and numerical processing interact in word problems in both adults and children population (Daroczy, Artemenko, et al., 2020; Daroczy, Meurers, et al., 2020). Specifically, linguistic features, including lexical consistency between relational terms and the required operation, and linguistic complexity such as nominalization, interact with arithmetic factors to shape how efficiently numerical relations are extracted and linked to an operation (Daroczy, Meurers, et al., 2020). This highlights that arithmetic complexity can be driven not only by numerical structure, but also by the contextual demands imposed by task format (paradigm) and problem framing (language).

Taken together, our study draws attention to paradigm as a context variable in arithmetic research. The limitations in our paradigm studies lie in the fact that strategy use was not directly measured, making it difficult to determine the reasons or underlying cognitive process for paradigm effects. Future work could therefore track the strategy or even manipulate the strategy, e.g., constrain or encourage specific strategies (Geurten & Lemaire, 2017; Tiberghien et al., 2017), to test whether paradigms still show differences under controlled conditions. As we discussed (previously and in Study 1 & 2) about the potential uncertainty in production paradigms and eliciting higher state math anxiety, math attitudes beyond state math anxiety, e.g., state math self-concept (Niepel et al., 2022), should also be included to test the proposed mechanisms more directly.

3.4 Math attitudes differences and dynamics

Beyond the arithmetic task itself and context variable of paradigm, individual differences also help explain why some arithmetic problems are harder than others and why some individuals struggle more. Although we have observed complexity effect at the group level in all the four studies in the dissertation, both interindividual and intraindividual differences should be acknowledged (Roth et al., 2024). As shown from the literature, carry and borrow operations involves place-value computation and integration, which needs both domain-general resources and domain-specific resources (Artemenko & Nuerk, 2025). Individuals may therefore differ in how efficiently they can meet these demands (e.g., processing speed and working memory capacity) and in how their subjective experience is (e.g., uncertainty and unconfident as well as anxiety).

The dissertation indicates that math attitudes are dynamic at both micro-timescale and macro timescale: beyond interindividual differences, math attitudes vary within individuals over task exposure and change with aging. In addition, trait math anxiety was reflected in reduced processing efficiency, whereas math self-concept was associated with a larger carry/borrow effect increase in error rates, indicating the importance of competence beliefs in coping with complex problems. Actually, math self-concept rather than anxiety was found to be more relevant for university major choices and math avoidance (Lunardon et al., 2025).

More specifically, math self-concept declines with age, while math anxiety increases, although performance in two-digit addition and subtraction with carry/borrow follows different patterns in terms of overall accuracy, which has higher error rates in childhood, decreasing by young adulthood and remaining relatively stable in middle and older age, while processing speed tends to decline with age (Avcil & Artemenko, 2025). This divergence between attitudes and performance suggests that underlying mechanisms follow different developmental trajectories: domain-general resources such as working memory show age-related declines but crystallized arithmetic knowledge and practiced multi-digit skills can compensate for these losses, at least up to a point. This also highlights that attitudes and performance are shaped by partly different forces: socio-psychological influences such as social comparison and evaluation pressure play a prominent role in math anxiety and self-concept (Beilock & Maloney, 2015; Marsh & Craven, 2006; Wigfield & Eccles, 2000), whereas age-related changes in arithmetic performance are more closely tied to cognitive ageing processes and compensatory mechanisms (Cabeza et al., 2018; Deary et al., 2009).

Taken together, these results underscore the need to consider both developmental stage and math attitudes when interpreting complexity effects in multi-digit arithmetic. A limitation here is that the lifespan patterns in our dissertation are based on cross-sectional data, which cannot fully separate aging effects from the generation and life experience differences (e.g., educational background).

3.5 Neural mechanisms

Across four studies in the dissertation, we observed complexity effect from both behavioral and neural level. Notably, the neural evidence and lifespan study in this dissertation are based on production paradigm, where individuals need to compute and generate the answer instead of choosing an answer from given choices. This methodological consistency makes the results comparable across studies but also constrains the interpretation of the observed neural patterns in different arithmetic contexts.

Study 4 used a production paradigm with fNIRS recording on three-digit addition and subtraction problems that varied carry and borrow demands (0, 1, or 2 carry/borrow operations), addressing the proposed question of whether increasing complexity mainly engages domain-general processes or also recruits domain-specific processes: Both domain-general executive regions and domain-specific numerical regions are increasingly recruited with complexity. Previous research also showed fronto-parietal shift in brain activation for arithmetic processing (Artemenko, 2021; Menon, 2010), indicating that children rely more on domain-general and less on domain-specific processes than adults, while adults were found to be more flexible in adopting different strategies for arithmetic (Moeller et al., 2011). Therefore, since the evidence about complexity neural mechanism is based on adult sample, whether the same neural patterns could be generalized to children or older adults across developmental stages remains to be further investigated (Vogel & De Smedt, 2021).

Moreover, the study revealed different modulation of these complexity effects by individual working memory capacity. We found that higher verbal short-term memory was associated with smaller complexity-related increases in bilateral MFG, whereas higher visuospatial short-term memory was associated with smaller increases in parietal regions (SMG/AG). However, because these working memory and brain activation relationships are correlational in this dissertation, future work could test causality by manipulating different working memory load (Imbo & LeFevre, 2010) and by directly assessing cognitive processing in each production tasks, e.g., using eye-tracking during producing answers (Goettfried et al., 2025) or trial-by-trial strategy reports (Grabner & De Smedt, 2011).

3.6 Practical implications

Focusing on arithmetic complexity, this dissertation examines paradigms and math attitudes as well as neural mechanisms for multi-digit arithmetic. The results and conclusions not only help us have a better understanding of arithmetic processing from theoretical level, but also provide practical implications for arithmetic fluency assessment, instructional design as well as interventions for math attitudes and cognition.

Implications for arithmetic fluency. The established arithmetic fluency tests often adopt production paradigms to test the mental calculation, and focus on speed and accuracy, e.g., how many correct answers in two minutes (Loenneker et al., 2024; Roy et al., 2025). Our results indicate that arithmetic performance is influenced not only by the task, but also by context (paradigm) as well as individual differences (math attitudes). Hence, in educational or clinical testing circumstances, cooperating decision paradigms for tasks as well as math attitudes including math anxiety and math self-concept could provide a comprehensive overview of arithmetic fluency.

Implications for instructional design. In classroom setting, teachers should choose task format based on their goals. Our findings suggest that paradigms can be used deliberately as instructional context or scaffolds. Decision paradigms provide answer options, which could serve as external cues for the calculation. Moreover, the well-designed distractors could be diagnostic and instructive: They can target common carry/borrow slips (e.g., omitting the carried or borrowed digit), which often yield off-by-ten errors in two-digit arithmetic, thereby helping students learn to detect and correct typical error types as well as informing teachers to develop target feedback and training (Arhin, 2024). In contrast, production paradigms are helpful in teaching and reinforcing computational algorithms as well as place-value understanding.

Implications for interventions. Our findings showed math attitudes and their dynamics during tasks as well as across lifespan. Accordingly, effective interventions should also be dynamic and making use of contexts. Since the state math anxiety shows highest level before the task begins, interventions could first target the anticipatory phase, e.g., emotional regulation or cognitive reappraisal before tasks (Sidney et al., 2025), and then applying graded exposure (Craske et al., 2014), starting with tasks in decision paradigms and gradually in production paradigms. Because complexity related error rates are more closely associated with math self-concept, interventions may be more effective when they go beyond anxiety reduction and actively strengthen perceptive competence. For example, integrating positive affirmations and teacher-reinforced growth mindset into daily classroom activities (Samuel & Warner, 2021). Given that math anxiety increases, and self-concept decreases through adulthood, such support should not be restricted to children and adolescents but should also be integrated into adult education and vocational training, where calculating or numeracy also plays an important role.

3.7 Conclusions

In summary, this dissertation clarifies why multi-digit arithmetic becomes difficult and why this difficulty does not look the same in everyone. Across four preregistered studies, the results conclude that complexity in multi-digit arithmetic is primarily driven by carry/borrow operations, which require place-value integration and increase demands on working memory. On behavioral level, these demands are expressed by slower reaction time. However, the effects on accuracy are sensitive to paradigm and individual differences.

This dissertation proposes that paradigms (production paradigm and decision decision) not only estimate arithmetic but also influence arithmetic performance and state math anxiety. This means that the results of studies conducted in different paradigms should be compared with caution, and paradigms should be considered as a context variable in future studies.

In addition, math attitudes of individuals are also linked to arithmetic performance and are dynamic both within task and across lifespan. It is worth noting that math self-concept, instead of math anxiety, is linked to error rates of complexity, which emphasizes the importance of perceived competence in dealing with complex tasks.

At neural level, the increasing complexity is associated with the recruitment of the fronto-parietal network, indicating the involvement of both domain-specific and domain-general resources. Additionally, working memory is correlated with smaller increases in activation related to complexity, indicating that cognitive resources can mitigate the neural expense of arithmetic.

Finally, all four studies were preregistered, and all materials, data, and analysis code were made publicly available, which promotes transparency in Open Science. In summary, this dissertation supports the taken home message: arithmetic complexity is not always equally difficult but also affected by context variable as well as individual differences.

References

Adams, G. S., Converse, B. A., Hales, A. H., & Klotz, L. E. (2021). People systematically overlook subtractive changes. *Nature*, *592*(7853), 258–261. https://doi.org/10.1038/s41586-021-03380-y

Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences*, *22*(3), 385–389. https://doi.org/10.1016/j.lindif.2011.12.004

Arhin, A. K. (2024). Developing Distractors for Mathematics Multiple Choice Items: A Literature Review. *Acta Educationis Generalis*, *14*(3), 103–120.

Arsalidou, M., & Taylor, M. J. (2011). Is 2 + 2 = 4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

Artemenko, C. (2018). *Neurocognitive Foundations of Arithmetic Complexity in Adults and Children*. Eberhard Karls Universität Tübingen.

Artemenko, C., Giannouli, V., & Nuerk, H.-C. (2024). Age-related effects in magnitude and place-value processing. *Scientific Reports*, *14*(1), 13645. https://doi.org/10.1038/s41598-024-63298-z

Artemenko, C., Masson, N., Georges, C., Nuerk, H.-C., & Cipora, K. (2021). Not all elementary school teachers are scared of math. *Journal of Numerical Cognition*, *7*(3), 275–294.

Artemenko, C., & Nuerk, H.-C. (2025). Empirical and conceptual relations of the carry effect in addition and the borrow effect in subtraction: An inversion account [Author Accepted Manuscript]. *Journal of Numerical Cognition*. https://www.psycharchives.org/en/item/aa152406-4046-44c6-8886-1af7d417badf

Artemenko, C., Soltanlou, M., Dresler, T., Ehlis, A.-C., & Nuerk, H.-C. (2018). The neural correlates of arithmetic difficulty depend on mathematical ability: Evidence from combined fNIRS and ERP. *Brain Structure and Function*, *223*(6), 2561–2574. https://doi.org/10.1007/s00429-018-1618-0

Artemenko, C., Soltanlou, M., Ehlis, A.-C., Nuerk, H.-C., & Dresler, T. (2018). The neural correlates of mental arithmetic in adolescents: A longitudinal fNIRS study. *Behavioral and Brain Functions*, *14*(1), 5. https://doi.org/10.1186/s12993-018-0137-8

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224. https://doi.org/10.1037/0096-3445.130.2.224

Ashcraft, M. H., & Stazyk, E. H. (1981). Menatal addition: A test of three verification models. *Memory & Cognition*, *9*(2), 185–196. https://doi.org/10.3758/BF03202334

Avcil, M., & Artemenko, C. (2025). Development of arithmetic across the lifespan: A registered report. *Developmental Psychology*. https://psycnet.apa.org/record/2026-00720-001

Bahnmueller, J., Nuerk, H.-C., & Moeller, K. (2018). A Taxonomy Proposal for Types of Interactions of Language and Place-Value Processing in Multi-Digit Numbers. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01024

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, *147*(2), 134–168. https://doi.org/10.1037/bul0000307

Beilock, S. L., & Maloney, E. A. (2015). Math Anxiety: A Factor in Math Achievement Not to Be Ignored. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 4–12. https://doi.org/10.1177/2372732215601438

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Brunner, M., Preckel, F., Götz, T., Lüdtke, O., & Keller, L. (2023). *The Relationship Between Math Anxiety and Math Achievement: New Perspectives From Combining Individual Participant Data and Aggregated Data in a Meta-Analysis*.

Campbell, J. I. D. (2008). Subtraction by addition. *Memory & Cognition*, *36*(6), 1094–1102. https://doi.org/10.3758/MC.36.6.1094

Campbell, J. I., & Fugelsang, J. (2001). Strategy choice for arithmetic verification: Effects of numerical surface form. *Cognition*, *80*(3), B21–B30. https://doi.org/10.1016/S0010-0277(01)00115-9

Caviola, S., Mammarella, I. C., Pastore, M., & LeFevre, J.-A. (2018). Children's Strategy Choices on Complex Subtraction Problems: Individual Differences and Developmental Changes. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01209

Cipora, K., Lunardon, M., Masson, N., Georges, C., Nuerk, H.-C., & Artemenko, C. (2024). The AMATUS Dataset: Arithmetic Performance, Mathematics Anxiety and Attitudes in Primary School Teachers and University Students. *Journal of Open Psychology Data*, *12*(1). https://doi.org/10.5334/jopd.115

Cipora, K., Santos, F. H., Kucian, K., & Dowker, A. (2022). Mathematics anxiety—Where are we and where shall we go? *Annals of the New York Academy of Sciences*, *1513*(1), 10–20. https://doi.org/10.1111/nyas.14770

Conlon, R. A., Hicks, A., Barroso, C., & Ganley, C. M. (2021). The effect of the timing of math anxiety measurement on math outcomes. *Learning and Individual Differences*, *86*, 101962. https://doi.org/10.1016/j.lindif.2020.101962

Craske, M. G., Treanor, M., Conway, C., Zbozinek, T., & Vervliet, B. (2014). Maximizing Exposure Therapy: An Inhibitory Learning Approach. *Behaviour Research and Therapy*, *58*, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Daroczy, G., Artemenko, C., Meurers, D., Wolska, M., & Nuerk, H.-C. (2020). Influence of Task Characteristics on Eye-Movement Patterns Related to Numerical and. *Neuroscience*, *9*, 1–8.

Daroczy, G., Meurers, D., Heller, J., Wolska, M., & Nürk, H.-C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing*, *21*(1), 105–125. https://doi.org/10.1007/s10339-019-00948-5

Daroczy, G., Soltanlou, M., Dresler, T., Artemenko, C., & Nuerk, H.-C. (2026). Language comprehension challenges arithmetic word problem solving – an fNIRS study. *ZDM – Mathematics Education*, 1–14. https://doi.org/10.1007/s11858-026-01765-9

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

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*(2), 219–250.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 626.

Dehaene, S., Piazza ,Manuela, Pinel ,Philippe, & and Cohen, L. (2003). Three Parietal Circuits for Number Processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Delazer, M., Domahs, F., Lochy, A., Karner, E., Benke, T., & Poewe, W. (2004). Number processing and basal ganglia dysfunction: A single case study. *Neuropsychologia*, *42*(8), 1050–1062.

Demedts, F., Reynvoet, B., Sasanguie, D., & Depaepe, F. (2022). Unraveling the role of math anxiety in students' math performance. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.979113

Dewi, J. D., Bagnoud, J., & Thevenot, C. (2021). Do production and verification tasks in arithmetic rely on the same cognitive mechanisms? A test using alphabet arithmetic. *Quarterly Journal of Experimental Psychology*, *74*(12), 2182–2192. https://doi.org/10.1177/17470218211022635

Ding, Y., Liu, R.-D., Liu, H., Wang, J., Zhen, R., & Jiang, R.-H. (2019). Effects of Working Memory, Strategy Use, and Single-Step Mental Addition on Multi-Step Mental Addition in

Chinese Elementary Students. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00148

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, 508.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Geurten, M., & Lemaire, P. (2017). Age-related differences in strategic monitoring during arithmetic problem solving. *Acta Psychologica*, *180*, 105–116. https://doi.org/10.1016/j.actpsy.2017.09.005

Goettfried, E., Zamarian, L., Goettfried, E., & Zamarian, L. (2025). Looking into the Calculating Mind: Evidence About Arithmetic from Eye-Tracking Studies. *Behavioral Sciences*, *15*(12). https://www.mdpi.com/2076-328X/15/12/1685

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & 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. https://doi.org/10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., & De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biological Psychology*, *87*(1), 128–136. https://doi.org/10.1016/j.biopsycho.2011.02.019

GrÉGoire, J., & Van Der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, *4*(2), 140–149. https://doi.org/10.1080/13825589708256642

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers in the Development of Gender-Related Math Attitudes. *Sex Roles*, *66*(3), 153–166. https://doi.org/10.1007/s11199-011-9996-2

Hickendorff, M. (2022). Flexibility and adaptivity in arithmetic strategy use: What children know and what they show. *Journal of Numerical Cognition*, *8*(3), 367–381.

Hinault, T., & Lemaire, P. (2016). What does EEG tell us about arithmetic strategies? A review. *International Journal of Psychophysiology*, *106*, 115–126. https://doi.org/10.1016/j.ijpsycho.2016.05.006

Huber, J. F., & Artemenko, C. (2021). Anxiety-Related Difficulties With Complex Arithmetic. *Zeitschrift Für Psychologie*.

Huber, S., Mann, A., Nuerk, H.-C., & Moeller, K. (2014). Cognitive control in number magnitude processing: Evidence from eye-tracking. *Psychological Research*, *78*(4), 539–548. https://doi.org/10.1007/s00426-013-0504-x

Huber, S., Nuerk, H.-C., Willmes, K., & Moeller, K. (2016). A general model framework for multisymbol number comparison. *Psychological Review*, *123*(6), 667–695. https://doi.org/10.1037/rev0000040

Imbo, I., & LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese-and Canadian-educated adults. *Memory and Cognition*, *38*(2), 176–185. https://doi.org/10.3758/MC.38.2.176

Imbo, I., Rammelaere, S. D., & Vandierendonck, A. (2005). New insights in the role of working memory in carry and borrow operations. *Psychologica Belgica*, *45*(2), Article 2. https://doi.org/10.5334/pb-45-2-101

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007a). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *Quarterly Journal of Experimental Psychology*, *60*(5), 708–731. https://doi.org/10.1080/17470210600762447

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological Research*, *71*(4), 467–483. https://doi.org/10.1007/s00426-006-0044-8

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007b). The role of working memory in carrying and borrowing. *Psychological Research*, *71*(4), 467–483. https://doi.org/10.1007/s00426-006-0044-8

Istomina, A., & Arsalidou, M. (2024). Add, subtract and multiply: Meta-analyses of brain correlates of arithmetic operations in children and adults. *Developmental Cognitive Neuroscience*, *69*, 101419. https://doi.org/10.1016/j.dcn.2024.101419

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children’s self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Jorm, A. F. (2000). Does old age reduce the risk of anxiety and depression? A review of epidemiological studies across the adult life span. *Psychological Medicine*, *30*(1), 11–22. https://doi.org/10.1017/S0033291799001452

Justicia-Galiano, M. J., Martín-Puga, M. E., Linares, R., & Pelegrina, S. (2017). Math anxiety and math performance in children: The mediating roles of working memory and math self-concept. *British Journal of Educational Psychology*, *87*(4), 573–589. https://doi.org/10.1111/bjep.12165

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The ‘Trier Social Stress Test’–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1–2), 76–81.

Klein, E., & Knops, A. (2023). The two-network framework of number processing: A step towards a better understanding of the neural origins of developmental dyscalculia. *Journal of Neural Transmission*, *130*(3), 253–268. https://doi.org/10.1007/s00702-022-02580-8

Klein, E., Moeller, K., Katharina Dressel, Domahs, F., Wood, G., Willmes, K., & Nuerk, H.-C. (2010). To carry or not to carry—Is this the question? Disentangling the carry effect in multi-digit addition. *Acta Psychologica*, *135*(1), 67–76.

Klein, E., Suchan, J., Moeller, K., Karnath, H.-O., Knops, A., Wood, G., Nuerk, H.-C., & Willmes, K. (2016). Considering structural connectivity in the triple code model of numerical cognition: Differential connectivity for magnitude processing and arithmetic facts. *Brain Structure and Function*, *221*(2), 979–995. https://doi.org/10.1007/s00429-014-0951-1

Klein, E., Willmes, K., Dressel, K., Domahs, F., Wood, G., Nuerk, H.-C., & Moeller, K. (2010). Categorical and continuous—Disentangling the neural correlates of the carry effect in multi-digit addition. *Behavioral and Brain Functions*, *6*(1), 70. https://doi.org/10.1186/1744-9081-6-70

Knops, A., Nuerk, H.-C., & Göbel, S. M. (2017). Domain-General Factors Influencing Numerical and Arithmetic Processing. *Journal of Numerical Cognition*, *3*(2), 112–132. https://doi.org/10.5964/jnc.v3i2.159

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research*, *22*(3), 397–405. https://doi.org/10.1016/j.cogbrainres.2004.09.011

Lambert, K., & Moeller, K. (2019). Place-value computation in children with mathematics difficulties. *Journal of Experimental Child Psychology*, *178*, 214–225. https://doi.org/10.1016/j.jecp.2018.09.008

Loenneker, H. D., Becker, S., Nussbaum, S., Nuerk, H.-C., & Liepelt-Scarfone, I. (2021). Arithmetic errors in financial contexts in Parkinson's disease. *Frontiers in Psychology*, *12*, 629984.

Loenneker, H. D., Cipora, K., Artemenko, C., Soltanlou, M., Bellon, E., De Smedt, B., García-Orza, J., Giannouli, V., Gutiérrez-Cordero, I., & Lipowska, K. (2024). Math4Speed: A freely available measure of arithmetic fluency. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*. https://psycnet.apa.org/record/2025-60418-001

Lunardon, M., Artemenko, C., Rossi, S., Nuerk, H.-C., & Cipora, K. (2025). More Than Just Anxiety: Math Attitudes as Key Driver of University Choice. *Annals of the New York Academy of Sciences*, *1553*, 240–255. https://doi.org/10.1111/nyas.70060

MacCormack, J. K., Bonar, A. S., & Lindquist, K. A. (2024). Interoceptive beliefs moderate the link between physiological and emotional arousal during an acute stressor. *Emotion*, *24*(1), 269.

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational Effects of Parents' Math Anxiety on Children's Math Achievement and Anxiety. *Psychological Science*, *26*(9), 1480–1488. https://doi.org/10.1177/0956797615592630

Marsh, H. W., & Craven, R. G. (2006). Reciprocal Effects of Self-Concept and Performance From a Multidimensional Perspective: Beyond Seductive Pleasure and Unidimensional Perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Menon, V. (2016). Memory and cognitive control circuits in mathematical cognition and learning. *Progress in Brain Research*, *227*, 159–186. https://doi.org/10.1016/bs.pbr.2016.04.026

Menon, V., Rivera, S., White, C., Eliez, S., Glover, G., & Reiss, A. (2000). Functional optimization of arithmetic processing in perfect performers. *Cognitive Brain Research*, *9*(3), 343–345.

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, *24*(Volume 24, 2001), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Moeller, K., Fischer, M. H., Nuerk, H.-C., & Willmes, K. (2009). Sequential or parallel decomposed processing of two-digit numbers? Evidence from eye-tracking. *Quarterly Journal of Experimental Psychology*, *62*(2), 323–334.

Moeller, K., Huber, S., Nuerk, H.-C., & Willmes, K. (2011). Two-digit number processing: Holistic, decomposed or hybrid? A computational modelling approach. *Psychological Research*, *75*(4), 290–306. https://doi.org/10.1007/s00426-010-0307-2

Moeller, K., Klein, E., & Nuerk, H.-C. (2011a). (No) small adults: Children's processing of carry addition problems. *Developmental Neuropsychology*, *36*(6), 702–720. https://doi.org/10.1080/87565641.2010.549880

Moeller, K., Klein, E., & Nuerk, H.-C. (2011b). Three processes underlying the carry effect in addition–Evidence from eye tracking. *British Journal of Psychology*, *102*(3), 623–645.

Moeller, K., Klein, E., Nuerk, H.-C., & Willmes, K. (2013). Magnitude representation in sequential comparison of two-digit numbers is not holistic either. *Cognitive Processing*, *14*(1), 51–62.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, *16*(4), 310–322. https://doi.org/10.1016/j.learninstruc.2006.07.003

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*(1), 185–208.

Niepel, C., Marsh, H. W., Guo, J., Pekrun, R., & Möller, J. (2022). Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson: An experience-sampling study. *Journal of Educational Psychology*, *114*(6), 1380.

Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the Mental Number Line. *Zeitschrift Für Psychologie*, *219*(1), 3–22. https://doi.org/10.1027/2151-2604/a000041

Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). Multi-digit Number Processing: Overview, conceptual clarifications, and language influences. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford Handbook of Numerical Cognition* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.021

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, *82*(1), B25–B33.

Nuerk, H.-C., Weger, U., & Willmes, K. (2004). On the perceptual generality of the unit-decade compatibility effect. *Experimental Psychology*, *51*(1), 72–79.

Nuerk, H.-C., & Willmes, K. (2005). On the magnitude representations of two-digit numbers. *Psychology Science*, *47*(1), 52–72.

Orbach, L., Herzog, M., & Fritz, A. (2019). Relation of state-and trait-math anxiety to intelligence, math achievement and learning motivation. *Journal of Numerical Cognition*, *5*(3), 371–399.

Passiatore, Y., Costa, S., Grossi, G., Carrus, G., & Pirchio, S. (2024). Mathematics self-concept moderates the relation between cognitive functions and mathematical skills in primary school children. *Social Psychology of Education*, *27*(3), 1143–1159. https://doi.org/10.1007/s11218-023-09854-3

Peters, G., De Smedt, B., Torbeyns, J., Ghesquière, P., & Verschaffel, L. (2010). Adults' use of subtraction by addition. *Acta Psychologica*, *135*(3), 323–329. https://doi.org/10.1016/j.actpsy.2010.08.007

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

Prado, J., Lu, J., Liu, L., Dong, Q., Zhou, X., & Booth, J. R. (2013). The neural bases of the multiplication problem-size effect across countries. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00189

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math Anxiety, Working Memory, and Math Achievement in Early Elementary School. *Journal of Cognition and Development*, *14*(2), 187–202. https://doi.org/10.1080/15248372.2012.664593

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2p1), 274.

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551–554. https://doi.org/10.1037/h0033456

Roos, A.-L., Bieg, M., Götz, T., Frenzel, A. C., Taxer, J., & Zeidner, M. (2015). Experiencing more mathematics anxiety than expected? Contrasting trait and state anxiety in high achieving students. *High Ability Studies*, *26*(2), 245–258.

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., & Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia*, *49*(9), 2592–2608. https://doi.org/10.1016/j.neuropsychologia.2011.04.035

Roth, L., Jordan, V., Schwarz, S., Willmes, K., Nuerk, H.-C., van Dijck, J.-P., & Cipora, K. (2024). Don’t SNARC me now! Intraindividual variability of cognitive phenomena – Insights from the Ironman paradigm. *Cognition*, *248*, 105781. https://doi.org/10.1016/j.cognition.2024.105781

Roy, E., Guillaume, M., Van Rinsveld, A., & McCandliss, B. D. (2025). Tablet-based arithmetic fluency assessment reveals developments in math cognition and math achievement from childhood to adolescence. *NPJ Science of Learning*, *10*, 19. https://doi.org/10.1038/s41539-025-00314-5

Samuel, T. S., & Warner, J. (2021). “I Can Math!”: Reducing Math Anxiety and Increasing Math Self-Efficacy Using a Mindfulness and Growth Mindset-Based Intervention in First-Year Students. *Community College Journal of Research and Practice*, *45*(3), 205–222. https://doi.org/10.1080/10668926.2019.1666063

Shomos, A., & Forbes, M. (2014). Literacy and numeracy skills and labour market outcomes in Australia. *Australian Government Productivity Commission*.

Sidney, P. G., Scheibe, D. A., Zahrn, L., Brown, K. G. I., & Thompson, C. A. (2025). Developing Effective Interventions for Math Anxiety. *Current Directions in Psychological Science*, *34*(1), 57–63. https://doi.org/10.1177/09637214241300111

Sokolowski, H. M., Hawes, Z., & Ansari, D. (2023). The neural correlates of retrieval and procedural strategies in mental arithmetic: A functional neuroimaging meta-analysis. *Human Brain Mapping*, *44*(1), 229–244.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, *101*(4), 862–882. https://doi.org/10.1037/a0024950

Spielberger, C. D. (2013). *Anxiety: Current trends in theory and research*. Elsevier.

Thevenot, C., Dewi, J. D. M., Bagnoud, J., Uittenhove, K., & Castel, C. (2020). Scrutinizing patterns of solution times in alphabet-arithmetic tasks favors counting over retrieval models. *Cognition*, *200*, 104272. https://doi.org/10.1016/j.cognition.2020.104272

Tiberghien, K., Notebaert, W., Smedt, B. D., & Fias, W. (2017). Reactive and Proactive Control in Arithmetical Strategy Selection. *Journal of Numerical Cognition*, *3*(3), 598–619. https://doi.org/10.5964/jnc.v3i3.124

Trbovich, P. L., & LeFevre, J.-A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, *31*(5), 738–745. https://doi.org/10.3758/BF03196112

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

Tushar, H., & Sooraksa, N. (2023). Global employability skills in the 21st century workplace: A semi-systematic literature review. *Heliyon*, *9*(11).

Viesel-Nordmeyer, N., & Prado, J. (2023). Arithmetic skills are associated with left fronto-temporal gray matter volume in 536 children and adolescents. *Npj Science of Learning*, *8*(1), 56.

Vogel, S. E., & De Smedt, B. (2021). Developmental brain dynamics of numerical and arithmetic abilities. *Npj Science of Learning*, *6*(1), 22. https://doi.org/10.1038/s41539-021-00099-3

Wang, Y. (2023). Self-concept, learning anxiety, and performance in mathematics learning: The moderating effect of teacher cognitive activation. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(9), em2323. https://doi.org/10.29333/ejmste/13499

Wen, R., & Dubé, A. K. (2022). A systematic review of secondary students' attitudes towards mathematics and its relations with mathematics achievement. *Journal of Numerical Cognition*, *8*(2), 295–325.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wortman, J., Lucas, R. E., & Donnellan, M. B. (2012). Stability and change in the Big Five personality domains: Evidence from a longitudinal study of Australians. *Psychology and Aging*, *27*(4), 867–874. https://doi.org/10.1037/a0029322

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

Yi-Rong, N., Si-Yun, S., Zhou-Yi, G., Si-Run, L., Yun, B., Song-Hao, L., & Chan, W. Y. (2011). Dissociated brain organization for two-digit addition and subtraction: An fMRI investigation. *Brain Research Bulletin*, *86*(5), 395–402. https://doi.org/10.1016/j.brainresbull.2011.08.016

Zbrodoff, N. J. (1999). Effects of counting in alphabet arithmetic: Opportunistic stopping and priming of intermediate steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 299.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General*, *115*(2), 118. https://doi.org/10.1037/0096-3445.115.2.118

Zbrodoff, N., & Logan, G. (1990). On the Relation Between Production and Verification Tasks in the Psychology of Simple Arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 83–97. https://doi.org/10.1037/0278-7393.16.1.83

Acknowledgements

I still remember the first day in Tübingen, stepped off bus No. 5 at Haydnweg, surrounded by beautiful falling leaves, and ate my first Brezel. Everything was new to me. Now, after learning so much about this small city, it seems about time to “get off the bus” once again and move on to the next stop. Looking back, I am grateful for what I gained here, in photos, in lunch breaks, in academic growth, and in the quieter changes within myself.

First and foremost, I want to thank my supervisors, Prof. Dr. Hans-Christoph Nürk and Dr. Christina Artemenko, for your invaluable guidance and patience. You both made me feel safe and supported whenever I ran into problems or questions, and your steady encouragement helped me grow with confidence. I also owe a special thank you to Prof. Yunfeng He, who acted as the bridge that brought me here to Tübingen.

Beyond academia, I feel incredibly lucky to have met such wonderful colleagues and visiting scholars at TÜBANG as well as my collaborator Dr. Beatrix Barth from Universitätsklinikum Tübingen. I still smile when I think of our Airbnb days together in Antwerp, London, and Loughborough for MCLS. Those moments of companionship have become precious memories, and I will carry them with me, perhaps even when I look back on Tübingen at eighty.

I am also deeply grateful to the China Scholarship Council for supporting my PhD journey at the University of Tübingen. This support made it possible for me to study in Germany and to meet so many international and Chinese friends along the way. At the same time, I would like to thank the LEAD graduate school for offering inspiring and wonderful retreats as well as exceptional research network. As a student at University of Tübingen, I have felt surrounded by kindness and humanistic care, which made both my life and study here very happy.

I would also like to thank my close friends back in China, especially those I stayed connected with online. Your support, conversations, and late-night discussions helped ease the loneliness of being far from home.

Finally, I want to express my deepest love and gratitude to my family, and especially to my dear sister. Thank you for your unwavering love, and for always seeing me and standing by me.

Appendix

A1. Study 1

A1.1 Publication of Study 1

Yao, X., Artemenko, C., He, Y., & Nuerk, H.-C. (2025). Arithmetic is not arithmetic: Paradigm matters for arithmetic effects. Cognition, 256, 106060. https://doi.org/10.1016/j.cognition.2024.106060.

Cognition 256 (2025) 106060

Contents lists available at ScienceDirect

Cognition

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

Full Length Article

Arithmetic is not arithmetic: Paradigm matters for arithmetic effects

Xinru Yao [a], Christina Artemenko [a], Yunfeng He [b,*], Hans-Christoph Nuerk [a,c]

[a] Department of Psychology, University of Tuebingen, Tuebingen, Germany
[b] Liaoning Key Laboratory of Psychological Testing and Behavior Analysis, Liaoning University, China
[c] German Centre for Mental Health (DZPG), Tuebingen, Germany

ARTICLE INFO

Keywords:
Two-digit arithmetic
Carry effect
Borrow effect
Paradigms
Production
Verification
Forced-choice

ABSTRACT

Research on arithmetic uses different experimental paradigms. So far, it is unclear whether these different paradigms lead to the same effects or comparable effect sizes. Therefore, this study explores how different experimental paradigms influence mental arithmetic performance, focusing on understanding the potential differences and similarities in cognitive processes between paradigms. Six paradigms were systematically compared: decision paradigms (verification, forced-choice, delayed forced-choice) and production paradigms (written production, verbal-keyboard production, and simple verbal production). The results show consistent arithmetic effects related to operation (addition vs. subtraction) and task difficulty (with or without carry/borrow) across all paradigms, particularly in reaction time measures. However, accuracy varied between paradigms, with verbal-keyboard production and simple verbal production paradigms showing higher effect sizes for accuracy measures. These findings underscore the importance of considering each paradigm's specific demands and characteristics in arithmetic research, suggesting that paradigm selection can influence the observed outcomes. Our study provides critical methodological insights that can guide future research in the design and interpretation of arithmetic tasks, enhancing the reliability and ecological validity of findings in numerical cognition.

1. Introduction

Every day humans need to calculate with multi-digit numbers (e.g., time, money). When practicing arithmetic in school, students typically receive arithmetic problems, calculate the answers, and then either say them out loud or write them down on paper. However, such naturalistic settings differ from highly controlled laboratory paradigms by multiple sources of information, hence resulting in oversimplified and too narrow theories with only limited implications for real-life mathematics development (Cantlon, 2020). This is not only theoretical speculation: A neuroimaging study in children has found that brain activation patterns during mathematics learning differed between naturalistic and controlled settings although there were some functional overlaps (Amalric & Cantlon, 2022; see also Artemenko, Soltanlou, Ehlis, et al., 2018). Thus, there is a need to systematically examine the impact and importance of the choice of a certain paradigm for research on arithmetic.

The paradigms used for research in arithmetic mostly differ from reality. On the one hand, neuroimaging studies on the neural correlates of arithmetic usually use verification or forced-choice paradigms due to methodological restrictions by fMRI or EEG (e.g., De Smedt et al., 2013; Grabner et al., 2009; Hinault & Lemaire, 2016; Tschentscher & Hauk, 2014). Furthermore, due to easier experimental control, even behavioral studies mostly use such paradigms, which can be answered by key presses (e.g., Campbell & Fugelsang, 2001; Thevenot et al., 2020; Zbrodoff, 1999). In verification or forced-choice paradigms, one or multiple answers are presented together with the arithmetic problem, and the participants decide whether a given answer is true or false (e.g., 26 + 53 = 69; verification paradigm) or which the correct answer is among multiple answers (e.g., 26 + 53 = 69 or 79; forced-choice paradigm). Because the decision regarding the given answer(s) is central in verification and forced-choice paradigms, we name them decision paradigms in the current study. On the other hand, the most frequently occurring paradigm for arithmetic tasks in school and even in daily life is production, i.e., children or adults are presented with an arithmetic problem and produce the answer themselves. In production paradigms, participants are required to solve the arithmetic problem mentally and respond by writing, typing, or speaking the answer (e.g., 26 + 53 =?; cf.

* Corresponding author at: Liaoning Key Laboratory of Psychological Testing and Behavior Analysis, Liaoning University, 66 Chongshan Road,110036 Shenyang, China.

E-mail address: heyf@lnu.edu.cn (Y. He).

https://doi.org/10.1016/j.cognition.2024.106060
Received 21 May 2024; Received in revised form 9 December 2024; Accepted 31 December 2024
Available online 10 January 2025

Artemenko, Soltanlou, Dresler, et al., 2018). Altogether, there is a gap between research, which uses mostly decision paradigms, and real-life arithmetic, with a larger ecological validity for production paradigms (Uittenhove & Lemaire, 2014).

Given the diversity of paradigms, an important question arises: Does the choice of paradigm significantly influence the reported arithmetic processes and effects? Some theories of arithmetic assume that verification, forced-choice, and production paradigms are essentially the same up to the end of the computation or retrieval stage and then begin to differ because of their response requirements (Ashcraft, 1984). According to this view, verification is essentially production plus comparison, with the correct answer being evaluated after its calculation (Ashcraft & Battaglia, 1978). An objection to these theories is that the equation in verification paradigms can be evaluated as a whole and participants can decide without computing or retrieving the true answer (Zbrodoff & Logan, 1986). Similar arguments have been made for forced-choice paradigms, where distractor effects imply that decisions may sometimes rely on rejecting incorrect answers rather than fully computing and verifying the correct result (e.g., Artemenko, Pixner, Moeller, & Nuerk, 2018; Moeller et al., 2011a). For instance, children typically search for the correct result (verifying the target) while adults sometimes stop if they find the distractor to be wrong (rejecting the distractor) (Moeller et al., 2011a). Further evidence against the hypothesis that verification is production plus comparison was found when imposing delays between arithmetic problems and answers. Such delays link verification and production paradigms, ranging from 0 ms (problem and answer appear simultaneously as in standard verification paradigm) to 1000 ms (sufficient time to compute or retrieve the true answer before the putative answer appears) (Zbrodoff & Logan, 1990). In this study, the presented answer influenced verification even when there's enough time to retrieve or calculate the solution. These findings are consistent with the idea that verification involves the evaluation of the equation as a whole and thus substantially differs from production (Campbell, 1987; Zbrodoff & Logan, 1990).

As different paradigms might yield both similar and different cognitive processes, performance might also differ between paradigms. For instance, two studies (Hecht, 2002; Seyler et al., 2003) found that working memory plays a minimal role in retrieval strategies, despite using different paradigms (verification/production) and operations (addition/subtraction). This suggests that different paradigms can yield sufficiently similar conclusions. However, Imbo and Vandierendonck (2007) challenged this conclusion, arguing that the paradigms and operations used in these studies were not directly comparable. By employing a production paradigm for both addition and subtraction, they found that the central executive was involved in both retrieval and procedural strategies, a finding that diverged from earlier studies. This underscores the importance of considering paradigm-specific effects when drawing conclusions in cognitive research.

Apart from paradigm differences, another critical aspect is the complexity of arithmetic tasks. Most research has focused on single-digit arithmetic, which may not generalize to more complex arithmetic involving multi-digit numbers (Nuerk et al., 2015) – typically needed in everyday life (e.g., during shopping). For multi-digit arithmetic, place-value computation plays an important role (e.g., calculations across units and decades). For instance, a carry operation in addition is required when the sum of the units of the operands exceeds 9, with a decade to be carried over (e.g., $36 + 28$ vs. $41 + 23$). Similarly, a borrow operation in subtraction is required whenever the unit of the subtrahend is larger than the unit of the minuend, and hence a decade has to be borrowed (e.g., 64–28 vs. 64–23). The carry effect in addition and the borrow effect in subtraction increase arithmetic difficulty in terms of response latencies and error rates (Artemenko, 2018; Artemenko, Soltanlou, Dresler, et al., 2018). This is because the carry and borrow operations particularly increase demands on working memory (Imbo et al., 2007; Imbo & LeFevre, 2010; Moeller et al., 2011b). Moreover, subtraction was found to be more difficult than addition and, likewise, the borrow effect was found to be larger than the carry effect (Artemenko, 2018).

While these effects have often been assumed to be independent of the paradigm used, focusing on task condition differences rather than overall accuracy or reaction time, only a few studies have considered paradigm specificity. For instance, Trbovich and LeFevre (2003) demonstrated that different presentation formats (e.g., horizontal vs. vertical problem presentation) affect arithmetic processing by influencing working memory demands. Effects of presentation format have also been found for basic numerical effects, such as the SNARC effect (Ito & Hatta, 2004), the compatibility effect (Pletzer et al., 2016), and number line estimation tasks (Dackermann et al., 2018). However, this research mainly focused on presentation format differences rather than paradigms, i.e., differences in response formats such as the presence or absence of presented answers. In our study, the presentation format will be the same across paradigms, allowing us to focus on how different cognitive processes are engaged depending on the paradigm.

To summarize, neurocognitive and behavioral studies use a variety of paradigms to investigate arithmetic effects and their underlying processes. Some paradigms have little ecological validity because, in school and everyday life, the correct answer typically has to be produced, not just verified or chosen. Moreover, the mathematical textbook analysis showed that children encounter more problems in production format (Siegler, 2024; Siegler & Oppenzato, 2021). This raises the question: Can the same effects be observed in different paradigms? And if so, are the effect sizes comparable between different paradigms? The problem is that if there are indeed strong paradigm-specific effects, it would challenge many conclusions derived from previous research on arithmetic. Conclusions about arithmetic processes as such might need to be narrowed down to a particular paradigm. On the other hand, if arithmetic effects are indeed consistent across paradigms, this would also be of utmost importance for future research: Researchers could select paradigms based on methodological (e.g., neuroimaging, online studies) or sample-specific considerations (e.g., children) without introducing paradigm-specific biases.[1]

To address this gap, our study aims to systematically investigate mental arithmetic across six different paradigms that are commonly used in research. Our goal is to identify the similarities and differences in arithmetic effects and performance across these paradigms. By doing so, we hope to provide a more nuanced understanding of how arithmetic processing is influenced by the specific paradigm used, thereby informing both theoretical models and practical applications in cognitive research. As preregistered (https://aspredicted.org/8sz6z.pdf), the following hypotheses on differences and similarities between paradigms are postulated:

Similarities and differences in arithmetic effects across paradigms. This study will investigate whether arithmetic effects vary across different paradigms. Specifically, it will examine whether the difficulty of arithmetic regarding the operation and carry/borrow effects differs depending on the paradigm. The following behavioral pattern of effects is expected regardless of the paradigm: (1) Subtraction is expected to be more difficult than addition (operation effect). (2) Difficulty of two-digit addition and subtraction increases whenever a carry or borrow operation is required (carry/borrow effects). (3) The borrow effect is expected to be larger than the carry effect (Artemenko, 2018). Even if the effects can be replicated in all paradigms, we will explore how the size and nature of these effects might vary depending on the paradigm. Furthermore, paradigm-specific effects, such as the distractor distance effect in the (delayed) forced-choice tasks, will be explored.

Differences in arithmetic performance across paradigms. Arithmetic

[1] Of course, it's always theoretically possible that different cognitive processes may underlie similar effects across various paradigms. However, the issue becomes more pressing when empirical research reveals that these effects and their magnitudes are already inconsistent across paradigms.

performance is expected to vary between decision paradigms and production paradigms due to differences in task demands: (1) Performance in decision paradigms is expected to be better than in production paradigms (production effect), as decision paradigms with given answers can sometimes be approached with estimation rather than full calculation. (2) Within decision paradigms, performance in the forced-choice paradigm is expected to be better than in the verification paradigm, as selecting between two given answers (one correct, one incorrect) might be easier than determining the correctness of a single answer. (3) The delayed forced-choice paradigm is expected to show slower reaction times compared to verification and forced-choice paradigms, as it might require some level of production before choosing an answer. (4) The written production paradigm is expected to show slower reaction times than the verbal production paradigms, due to the complexity of the response format.

2. Methods

2.1. Participants

In total, 65 participants (17 male, 46 female, 2 diverse nonbinary; age: $M = 22.86$ years, $SD = 3.80$ years) participated in the study, one participant dropped out. Among all participants, 57 were right-handed and 8 were left-handed. Inclusion criteria for participants included being aged between 18 and 40 years old, being native German speakers, and having no dyscalculia or other learning disorders (e.g., attention deficit hyperactivity disorder). For participation, all participants received student credits or monetary reimbursement. Informed written consent was obtained from all participants and the study was conducted conforming to the latest version of the Declaration of Helsinki.

2.2. Materials

The mental arithmetic task was used in all six paradigms. Each arithmetic problem consisted of two two-digit operands that resulted in a two-digit answer. The problems included addition with (e.g., 36 + 27) or without (e.g., 32 + 24) carrying as well as subtraction with (e.g., 63–25) or without (e.g., 69–23) borrowing in a 2 operation (addition/subtraction) × 2 difficulty (simple/complex) design. Each of these four conditions consisted of 24 problems resulting in 96 problems within one stimulus set.

For the six paradigms, twelve matched stimulus sets (including six simple datasets and six complex datasets) were created. Note that the stimulus sets were matched, but not identical to avoid trial-specific learning across paradigms. Within and across all stimulus sets, the addition problems were matched in the numerical magnitude of the operands and problem size as the sum of the operands (see Supplementary Material, Table S7). The position of the larger operand was counterbalanced within each condition and set. In order to avoid the case of easier-solving problems, the stimulus sets did not include pure decades (e.g., 20) and ties (e.g., 22) as operands or answers nor any unit ties (e.g., 32 + 12) and decade ties (e.g., 21 + 23) between the operands (cf. Artemenko, Soltanlou, Dresler, et al., 2018; Klein et al., 2009; LeFevre et al., 2004; Nuerk et al., 2002). The subtraction problems were constructed as the inverse of the addition problems (e.g., 25 + 31 → 56–31), but they were not presented together, neither within the same paradigm nor in the nearby paradigms. All stimuli and paradigms are available as Open Material (https://osf.io/z8cqm/).

The arithmetic task was presented in six different experimental paradigms (see Fig. 1):

i. *Verification*. In the verification paradigm, an equation consisting of the arithmetic problem, an equal sign, and an answer was presented, and the participants were asked to indicate whether it was true or false by a button press of the left or right marked key with the index finger of the left or right hand, respectively. In this paradigm, half of the problems were true equations, and the rest were false equations with a distractor instead of the true answer.
ii. *Forced-choice*. In the forced-choice paradigm, an arithmetic problem was simultaneously presented with two answers below

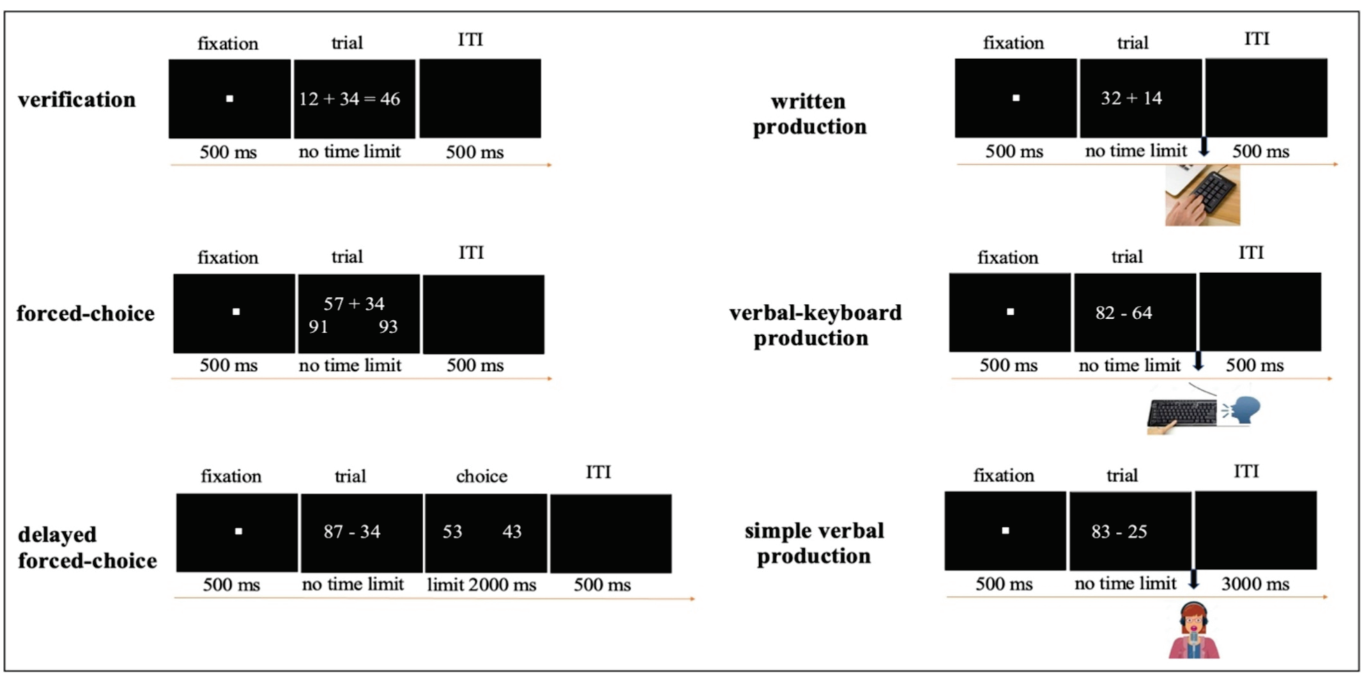


Fig. 1. Different arithmetic paradigms.
Notes. Verification means participants need to indicate whether the given answer is right or wrong. Forced-choice means participants need to choose one out of two given answers. Delayed forced-choice means choosing one out of two answers after participants already have an answer in mind. Written production means typing out the answer directly in a number keyboard. Verbal-keyboard production means participants speak out the answer while pressing a button on the keyboard. Simple verbal production means participants speak out the answer directly.

(target and distractor), and the participants were asked to decide whether the left or the right answer was correct by button press of the left or right marked key, respectively.

iii. *Delayed forced-choice.* In the delayed forced-choice paradigm, an arithmetic problem was presented first, and the participants were asked to indicate by pressing the space bar when they have come to an answer for that problem. Then two answers (target and distractor, without the arithmetic problem) were presented, and the participants were asked to decide whether the left or the right answer is correct by button press of the left or right marked key, respectively.

iv. *Written production.* In the written production paradigm, an arithmetic problem was presented, and the participants were asked to type in the answers with a number keyboard.

v. *Verbal-keyboard production.* In the verbal-keyboard production paradigm, an arithmetic problem was presented, and the participants were instructed to speak the answer out loud while holding down the space bar on the keyboard until they finished speaking.

vi. *Simple verbal production.* In the simple verbal production paradigm, an arithmetic problem was presented, and the participants were asked to speak out the answer directly.

The verification, forced-choice, and delayed forced-choice paradigms further included distractors with a distance from the true answer of ± 2 (different units; e.g., $37 + 25 = 64$ or 62) and $\pm$ 10 (different decades; e.g., $37 + 25 = 52$ or 62) in equal parts to ensure that both decades and units are processed in two-digit arithmetic (cf. Artemenko et al., 2015).

For the arithmetic task, split-half reliabilities (corrected for attenuation due to half length) were calculated separately for each paradigm and condition (see Supplementary Materials, Table S8a). The split-half reliabilities for reaction time (RT) were high in production paradigms (0.75–0.97), particularly in the verbal-keyboard production paradigm ($\geq$ 0.93), and lower but still acceptable in decision paradigms (0.61–0.86; only one exception of 0.24). The split-half reliabilities for accuracy (ACC) were rather low in decision paradigms (probably due to ceiling effects) and the written production paradigm ($\leq$ 0.41), and only acceptable for the verbal production paradigms (0.60–0.82; only one exception with 0.25). Additionally, construct validity was determined by correlating performance between the different paradigms (see Supplementary Materials, Tables S8b & S8c), resulting in significant correlations for RT (0.62–0.82) and ACC (0.30–0.73).

2.3. Procedure

The arithmetic task in different paradigms (see Fig. 1 for a flowchart of the six paradigms) was computerized using the software OpenSesame 3.3.10 (Mathôt et al., 2012). The arithmetic problems were presented in white ink in the center of a black screen. Response keys on a QWERT keyboard were "D" marked with a yellow dot on the left, "K" marked with a red dot on the right, and the space bar. In the written production paradigm, we used an extra number keyboard for typing the digits. Each trial in all paradigms started with a fixation dot for 500 ms, then the arithmetic problem was presented without a time limit until the participants' response. To emphasize mental arithmetic, the arithmetic problem immediately disappeared when the response started. Only in the delayed forced-choice paradigm the answers were presented separately until the participants' response with a time limit of 2000 ms. Each trial ended with a black screen for 500 ms after the participants' response, except for the simple verbal production paradigm, where 3000 ms was used to provide enough time for oral responses. The dependent variables were reaction time (RT) and accuracy (ACC). RTs were defined from stimulus onset until the first button press or voice key response; the sound threshold parameter was set to 0.06 (with normalized unit) to define the minimum amplitude level for significant auditory signal detection. For the reaction time of delayed forced-choice paradigm, an additional second RT was defined as starting from the first button press to indicating the choice with the second button press.

The order of the paradigms followed a Latin Square design with alternating decision and production paradigms, which ensured that each paradigm appeared in each ordinal position (e.g., first, second, third…) equally across participants to mitigate potential order effects. The Latin Square design controls for the position of each condition but does not fully counterbalance the relative order of conditions across participants. The order of paradigms was verification, verbal-keyboard production, forced-choice, written production, delayed forced-choice, and simple verbal production. Each of the six sequences started with another paradigm. Participants were instructed to solve the arithmetic problems as accurately and quickly as possible. Test anxiety, math anxiety, math self-concept, state anxiety, state math anxiety, and related constructs were also assessed for other reasons but were not part of the current study.

2.4. Analysis

Participants were removed (case-wise exclusion per paradigm; see Table S1 for details) if they had missing data, an ACC below 50 % per production paradigm or below 75 % per decision paradigm (due to a 50 % chance level), or a mean RT more than 3 median absolute deviations (MAD; Bayot et al., 2018) above or below the group *Median* for the respective paradigm. Trials were removed from RT analysis according to the following criteria (see Table S1 for details): false equation trials in the verification paradigm, incorrectly solved trials (i.e., errors and missings), RTs below 200 ms (anticipations), RTs more than 3 *MAD* above or below the individual *Median* for the respective paradigm (outliers), or distance between first and second RT more than 3 *MAD* above or below the individual *Median* for the respective paradigm (delayed forced-chocie, written production, and verbal-keyboard production). In addition, a logit transformation was applied to ACC data; no transformation was applied to RT data. Consequently, RT analysis was based on correct trials only and cleaned for outliers.

The preregistered analyses (https://aspredicted.org/8sz6z.pdf) were conducted using JASP (Jeffreys's Amazing Statistics Program, Version 0.18, JASP Team, 2023) (http://www.jasp-stats.org) using frequentist and Bayesian statistics to evaluate evidence for an effect or evidence for a null effect. The Bayes factor (*BF*) is defined as how much more likely the observed data is under one compared to the other hypothesis. BF_{10} means a Bayes factor in favor of the alternative hypothesis (evidence for a difference when $BF_{10} > 1$) and BF_{01} means a Bayes factor in favor of the null hypothesis (evidence for no difference when $BF_{01} > 1$), whereby $BF_{01} = 1/BF_{10}$ (Faulkenberry et al., 2020). A *BF* between 1 and 3 indicates anecdotal evidence, a *BF* between 3 and 10 moderate evidence, a *BF* between 10 and 30 indicates strong evidence, a *BF* between 30 and 100 very strong evidence, and a *BF* above 100 extreme evidence in favor of one hypothesis (Jeffreys, 1998; Lee & Wagenmakers, 2014).

First, the statistical analyses of RT and ACC were performed in 6 paradigm (verification, forced-choice, delayed forced-choice, written production, verbal-keyboard production, simple verbal production) $\times$ 2 operation (addition, subtraction) $\times$ 2 difficulty (simple, complex) repeated-measures ANOVAs and Bayesian ANOVAs to investigate arithmetic performance dependent on paradigms. Second, 2 operation (addition, subtraction) $\times$ 2 difficulty (simple, complex) repeated-measures ANOVAs and Bayesian ANOVAs on RT and ACC were separately conducted for each verification and production paradigm to explore specific arithmetic effects. Likewise, 2 operation (addition, subtraction) $\times$ 2 difficulty (simple, complex) $\times$ 2 distractor distance (2,10) repeated-measures ANOVAs and Bayesian ANOVAs were conducted on ACC and RT. Following the ANOVA, pairwise comparisons between conditions were Bonferroni-Holm corrected. In an exploratory analysis, the same analysis was conducted with the second RT in the delayed forced-choice paradigm.

3. Results

3.1. Arithmetic performance differs among paradigms

The three-way ANOVA and Bayesian ANOVA showed strong evidence for main effects of paradigm, operation, and difficulty on both accuracy and reaction time (see Table 1). How arithmetic performance differs between the six paradigms is presented in Fig. 2 (for descriptive data see also Supplementary Materials, Table S2) and will be elaborated in the following.

Differences between paradigms. Arithmetic performance differed depending on the paradigm, in the following way: (1) Decision paradigms were more accurate than production paradigms. (2) Within production paradigms, verbal-keyboard production was both more accurate and faster than written production, and also more accurate (while equally fast) than simple verbal production. (3) Forced-choice was faster than verification and production, while equally accurate compared to the other two decision paradigms. (4) Written production was the slowest of all paradigms and also less accurate, except for simple verbal production. In the following, these points will be explained in detail.

(1) Convergent evidence showed that accuracy in decision paradigms (verification, forced-choice, and delayed forced-choice) was respectively higher than in production paradigms (written production, simple verbal production), indicating that the presence of pre-provided correct or incorrect answers in decision paradigms facilitates the mental arithmetic solving process. Moreover, the forced-choice paradigm also was faster than the production paradigms, refuting a potential speed-accuracy trade-off in this paradigm; however, for the other two decision paradigms this was only the case in comparison to the written production paradigm. The verbal-keyboard production paradigm differed in accuracy and reaction time only significantly from the forced-choice paradigm without a speed-accuracy trade-off, while performance in this paradigm was rather similar to the other decision paradigms.

(2) Accuracy did not significantly differ between the different production paradigms, but Bayesian evidence pointed at higher accuracy in verbal-keyboard production than in written production and simple verbal production. This difference in production response formats indicates that additionally pressing a button on the keyboard might be more accurate than typing the answer or speaking out the answer directly. As response time was similar, the only difference between the two verbal production paradigms was whether a button was pressed or the voice key was activated, suggesting two possible explanations for the accuracy difference: On the one hand, despite using a robust threshold to detect voice onset, occasional interference from uncontrolled environmental noise activated the voice key before the participant could calculate the answer. In such cases, missing trials were counted as incorrectly solved trials (as preregistered) and as such contributed to the lower accuracy in the simple verbal production paradigm. Our exploratory analysis showed that approximately 5 % of trials in the simple verbal production paradigm were missing, compared to nearly 0 % in the verbal-keyboard production paradigm, highlighting the impact of these technical issues. For reaction times, such trials were excluded from analysis (as preregistered) and thus the reaction time results were less affected by this technical issue. On the other hand, embodiment congruent gestures or direct touch were found to promote performance in math (Carlson et al., 2007; Segal, 2011). Thus, the button press while answering might have provided a buffer to better compute and think in the verbal-keyboard production paradigm.

(3) The verification paradigm was significantly slower than the forced-choice paradigm, despite similar accuracy. Besides, the hard forced-choice items (+/− 10) are not consistently easier than verification, while the easier forced-choice items (+/− 2) are consistently easier than verification (see Supplementary Material, Table S9a & S9b). In verification paradigm, only one answer was presented, but in forced-choice paradigm, two answers are presented, one of them being the distractor. We intentionally employed the values of ± 2 and $\pm$ 10 for the distance of the distractor to the target, so that relying solely on either the unit or the decade proves insufficient. This deliberate design eliminates simple shortcuts and heightens task complexity compared to strategies involving a singular focus on the unit. In forced-choice paradigms, adults use two strategies: verifying the correct result and rejecting the distractor (Moeller et al., 2011a). Thus, better performance in the forced-choice paradigm may therefore be due to flexible strategy choice.

Bayesian evidence also showed this accelerated solution process in the forced-choice paradigm also in comparison to the production paradigms. This is essential for our understanding of mental arithmetic: In production paradigms, we measure calculation processes, while in the forced-choice paradigm, we measure a combination of calculation and distractor rejection processes. It can be imagined much like in a multiple-choice task in exams or quiz shows: You do not need to know the correct answer to be correct – you can also just exclude incorrect

Table 1
Results for arithmetic performance dependent on paradigm, operation, and difficulty.

		ANOVA			*Bayesian ANOVA*			
DV	factor	*F*	*p*	η_p^2	P_{incl}	$P_{incl\|data}$	BF_{incl}	BF_{excl}
ACC	paradigm	11.64	**< 0.001**	0.18	0.26	0.98	**> 100**	0.00
	operation	14.97	**< 0.001**	0.22	0.26	0.30	**27.46**	0.04
	difficulty	37.95	**< 0.001**	0.41	0.26	0.31	**> 100**	0.00
	paradigm × operation	1.07	0.376	0.02	0.26	0.02	0.02	**67.13**
	paradigm × difficulty	0.85	0.516	0.02	0.26	0.01	0.01	**95.39**
	operation × difficulty	8.50	**0.005**	0.14	0.26	0.69	2.27	0.44
	paradigm × operation × difficulty	0.81	0.540	0.02	0.05	0.00	0.03	**32.52**
RT	paradigm	44.86	**< 0.001**	0.45	0.26	0.00	**> 100**	0.00
	operation	177.92	**< 0.001**	0.77	0.26	0.72	**> 100**	0.00
	difficulty	417.15	**< 0.001**	0.89	0.26	0.00	**> 100**	0.00
	paradigm × operation	2.01	0.077	0.04	0.26	0.11	0.13	**7.85**
	paradigm × difficulty	11.06	**< 0.001**	0.17	0.26	1.00	**> 100**	0.00
	operation × difficulty	0.66	0.421	0.01	0.26	0.19	0.23	**4.31**
	paradigm × operation × difficulty	0.67	0.644	0.01	0.05	0.00	0.02	**63.20**

Notes. Pairwise comparion results can be found in the supplementary materials (Table S3a & S3b). P_{incl} represents the prior inclusion probability of an effect, $P_{incl|data}$ represents the inclusion probability given the data. BF_{incl} stands for the Bayes Factor for inclusion, indicating evidence for a difference; BF_{excl} stands for the Bayes factor for exclusion, indicating evidence for a null effect.

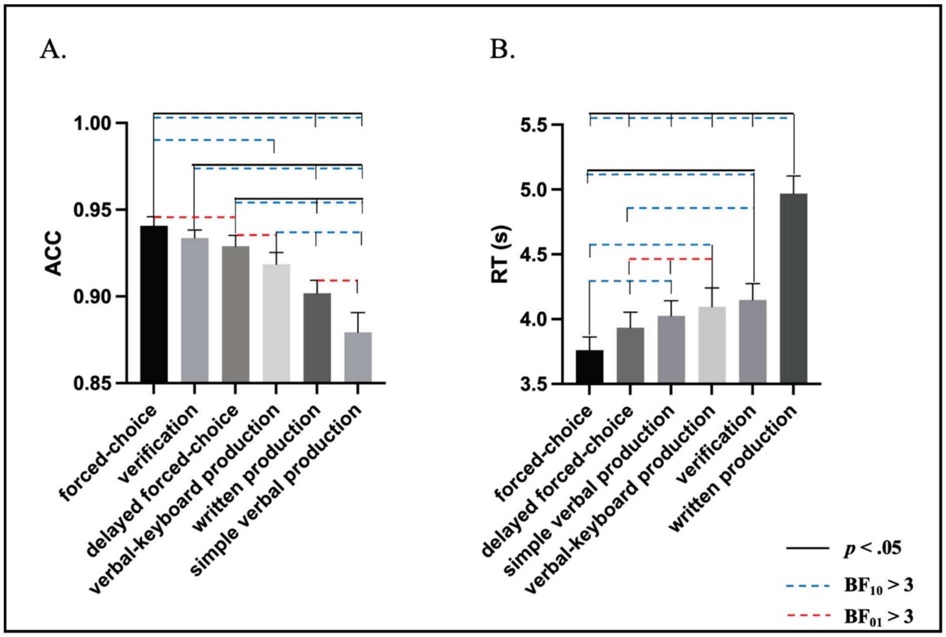


Fig. 2. Arithmetic performance across six paradigms.
Notes. A. Accuracy across six paradigms; B. Reaction time (s) across six paradigms. Error bars represent the standard error of the mean. Solid lines indicate significant differences from frequentist analysis and dashed lines indicate evidence for differences (blue) or evidence for null effects (red) from Bayesian analysis. Note that for reasons of accessibility, the order of paradigms in the Figure is according to performance, which slightly differs between ACC and RT. For accuracy, raw data is presented, while statistical analysis is based on logit-transformed accuracy data.

answers. Therefore, the prevalent use of decision paradigms in research may artificially boost arithmetic performance and potentially inflate estimates of arithmetic proficiency.

(4) The written production paradigm was significantly slower than all other paradigms, while being also less accurate (except for simple verbal production). Without a speed-accuracy trade-off, this means worse performance in written production paradigm, probably due to the response format requiring motor planning and execution for typing the answer. Besides these broader domain-general requirements (unrelated to numerical processing), the potential influence of specific numerical factors might also be considered, as another effect in two-digit number processing, which was found in verbal production, disappeared when manually typing the response (Brysbaert et al., 1998). This implies that written production could prompt alternative calculation strategies, emphasizing decomposed place-value processing over natural number naming.

3.2. Arithmetic effects across paradigms

The arithmetic effects (operation effect, difficulty effect, and distractor effect) are summarized for the six paradigms in Table 2. All observed arithmetic effects were in the same direction.

Operation effects. Paradigm-specific ANOVAs and Bayesian ANOVAs revealed a significant operation effect in all six paradigms for reaction time, but only in the two verbal production paradigms for accuracy (see Tables 2, 3, 4; Fig. 3; Supplementary Materials, Tables S4a & S4b). This replicates previous research suggesting that subtraction is generally more cognitively demanding than addition, possibly due to additional steps required in subtraction leading to higher reaction times and lower accuracy (Artemenko, Pixner, Moeller, & Nuerk, 2018; Imbo et al., 2007). While reaction time and accuracy were consistent for the operation effect in the two verbal production paradigms, the operation effect was not found for accuracy in the other paradigms, particularly the decision paradigms (probably due to ceiling effects). In summary, the operation effect in reaction time was independent of the paradigm, while the operation effect in accuracy depent on the paradigm. (See

Table 2
Summary of arithmetic effects in six paradigms for accuracy and reaction time.

		decision paradigms			production paradigms		
		verification	forced-choice	delayed forced-choice	written production	verbal-keyboard production	simple verbal production
operation effect	ACC	×	×	×	×	✓	✓
	RT	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
difficulty effect	ACC	×	✓	✓	✓	✓✓	✓
	RT	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓
interaction effect (operation × difficulty)	ACC	×	×	×	×	×	×
	RT	×	×	×	×	×	×
distractor effect	ACC	–	✓✓✓	✓	–	–	–
	RT	–	✓✓✓	× | ✓	–	–	–

Notes. "×" means no effect; "✓" means small effect ($|d| < 0.5$), "✓✓" means medium effect ($0.5 \leq |d| < 0.8$), "✓✓✓" means large effect ($|d| \geq 0.8$); "-"means irrelevant to this paradigm, × | ✓ means that there was no distractor effect for the first reaction time (calculation phase) while there was a small distractor effect for the second reaction time (response phase). Please note that the null effects of operation and difficulty in decision paradigms could be due to a ceiling effect.

Table 3
Results of paradigm-specific analyses of accuracy.

		ANOVA			Bayesian ANOVA			
paradigm	factor	F	p	η^2_p	P_{incl}	$P_{incl\|data}$	BF_{incl}	BF_{excl}
	operation	0.58	0.448	0.01	0.40	0.16	0.20	**5.02**
	difficulty	3.82	0.055	0.06	0.40	0.39	0.69	1.46
verification	operation × difficulty	1.86	0.178	0.03	0.20	0.04	0.67	1.50
	operation	3.70	0.059	0.06	0.26	0.29	0.62	1.62
	difficulty	15.14	**< 0.001**	0.21	0.26	0.57	**65.68**	0.02
	distractor	61.34	**< 0.001**	0.51	0.26	0.63	**> 100**	0.00
	operation × difficulty	3.19	0.080	0.052	0.26	0.17	0.49	2.05
	operation × distractor	0.17	0.679	0.00	0.26	0.09	0.21	**4.80**
	difficulty × distractor	2.05	0.158	0.03	0.26	0.30	0.44	2.26
forced-choice	operation × difficulty × distractor	0.54	0.466	0.01	0.05	0.00	0.26	**3.80**
	operation	2.52	0.118	0.04	0.26	0.19	0.28	**3.64**
	difficulty	8.11	**0.006**	0.12	0.26	0.69	**4.73**	0.21
	distractor	12.78	**< 0.001**	0.17	0.26	0.76	**24.08**	0.04
	operation × difficulty	0.57	0.454	0.01	0.26	0.05	0.20	**4.93**
	operation × distractor	2.29	0.136	0.04	0.26	0.10	0.45	2.22
	difficulty × distractor	0.29	0.593	0.01	0.26	0.13	0.18	**5.55**
delayed forced-choice	operation × difficulty × distractor	0.22	0.640	0.00	0.05	0.00	0.21	**4.80**
	operation	2.29	0.136	0.04	0.40	0.28	0.46	2.19
	difficulty	10.90	**0.002**	0.15	0.40	0.79	**7.21**	0.14
written production	operation × difficulty	1.52	0.222	0.02	0.20	0.10	0.42	2.37
	operation	11.39	**0.001**	0.15	0.40	0.80	**15.07**	0.07
	difficulty	29.47	**< 0.001**	0.32	0.40	0.85	**> 100**	0.00
verbal-keyboar production	operation × difficulty	0.15	0.702	0.00	0.20	0.15	0.19	**5.29**
	operation	9.61	**0.003**	0.13	0.40	0.46	**8.92**	0.11
	difficulty	24.80	**< 0.001**	0.29	0.40	0.51	**> 100**	0.00
simple verbal production	operation × difficulty	3.49	0.066	0.05	0.20	0.49	1.08	0.93

Notes. Pairwise comparion results can be found in the supplementary materials (Table S4a & S5). P_{incl} represents the prior inclusion probability of an effect, $P_{incl|data}$ represents the inclusion probability given the data. BF_{incl} stands for the Bayes Factor for inclusion, indicating evidence for a difference; BF_{excl} stands for the Bayes factor for exclusion, indicating evidence for a null effect.

Table 4
Results of paradigm-specific analyses of reaction time.

		ANOVA			Bayesian ANOVA			
paradigm	factor	F	p	η^2_p	P_{incl}	$P_{incl\|data}$	BF_{incl}	BF_{excl}
	operation	109.15	**< 0.001**	0.65	0.40	0.71	**> 100**	0.00
	difficulty	184.91	**< 0.001**	0.76	0.40	0.71	**> 100**	0.00
verification	operation × difficulty	1.66	0.203	0.03	0.20	0.29	0.40	2.47
	operation	97.83	**< 0.001**	0.63	0.26	0.00	**> 100**	0.00
	difficulty	243.16	**< 0.001**	0.81	0.26	0.57	**> 100**	0.00
	distractor	77.97	**< 0.001**	0.57	0.26	0.00	**> 100**	0.00
	operation × difficulty	0.17	0.686	0.00	0.26	0.19	0.24	**4.09**
	operation × distractor	17.75	**< 0.001**	0.23	0.26	0.98	**> 100**	0.01
	difficulty × distractor	1.67	0.201	0.03	0.26	0.28	0.39	2.54
forced-choice	operation × difficulty × distractor	0.52	0.475	0.01	0.05	0.01	0.23	**4.35**
	operation	79.16	**< 0.001**	0.56	0.26	0.64	**> 100**	0.00
	difficulty	251.02	**< 0.001**	0.80	0.26	0.61	**> 100**	0.00
	distractor	1.33	0.254	0.02	0.26	0.16	0.21	**4.69**
	operation × difficulty	2.13	0.150	0.03	0.26	0.34	0.52	1.94
	operation × distractor	0.03	0.866	0.00	0.26	0.04	0.16	**6.25**
	difficulty × distractor	2.26	0.138	0.04	0.26	0.07	0.40	2.51
delayed forced-choice	operation × difficulty × distractor	0.12	0.736	0.00	0.05	0.00	0.19	**5.32**
	operation	91.01	**< 0.001**	0.60	0.40	0.82	**> 100**	0.00
	difficulty	303.23	**< 0.001**	0.83	0.40	0.82	**> 100**	0.00
written production	operation × difficulty	0.45	0.504	0.01	0.20	0.18	0.22	**4.58**
	operation	136.18	**< 0.001**	0.68	0.40	0.39	**> 100**	0.00
	difficulty	331.81	**< 0.001**	0.84	0.40	0.39	**> 100**	0.00
verbal-keyboard production	operation × difficulty	5.00	**0.029**	0.07	0.20	0.61	1.59	0.63
	operation	150.39	**< 0.001**	0.71	0.40	0.83	**> 100**	0.00
	difficulty	295.48	**< 0.001**	0.83	0.40	0.83	**> 100**	0.00
simple verbal production	operation × difficulty	0.30	0.584	0.01	0.20	0.17	0.20	**4.91**

Notes. Pairwise comparion results can be found in the supplementary materials (Table S4b & S5). P_{incl} represents the prior inclusion probability of an effect, $P_{incl|data}$ represents the inclusion probability given the data. BF_{incl} stands for the Bayes Factor for inclusion, indicating evidence for a difference; BF_{excl} stands for the Bayes factor for exclusion, indicating evidence for a null effect.

Table 5.)

Difficulty effects. Paradigm-specific ANOVAs and Bayesian ANOVAs revealed difficulty effects for both reaction time and accuracy (except for the verification paradigm) in all six paradigms (see Tables 2, 3 4; Fig. 3; Supplementary Materials, Tables S4a & S4b). This replicates the carry and borrow effects indicating that addition with carrying and subtraction with borrowing are slower and less accurately solved than addition without carrying and subtraction without borrowing. Hence,

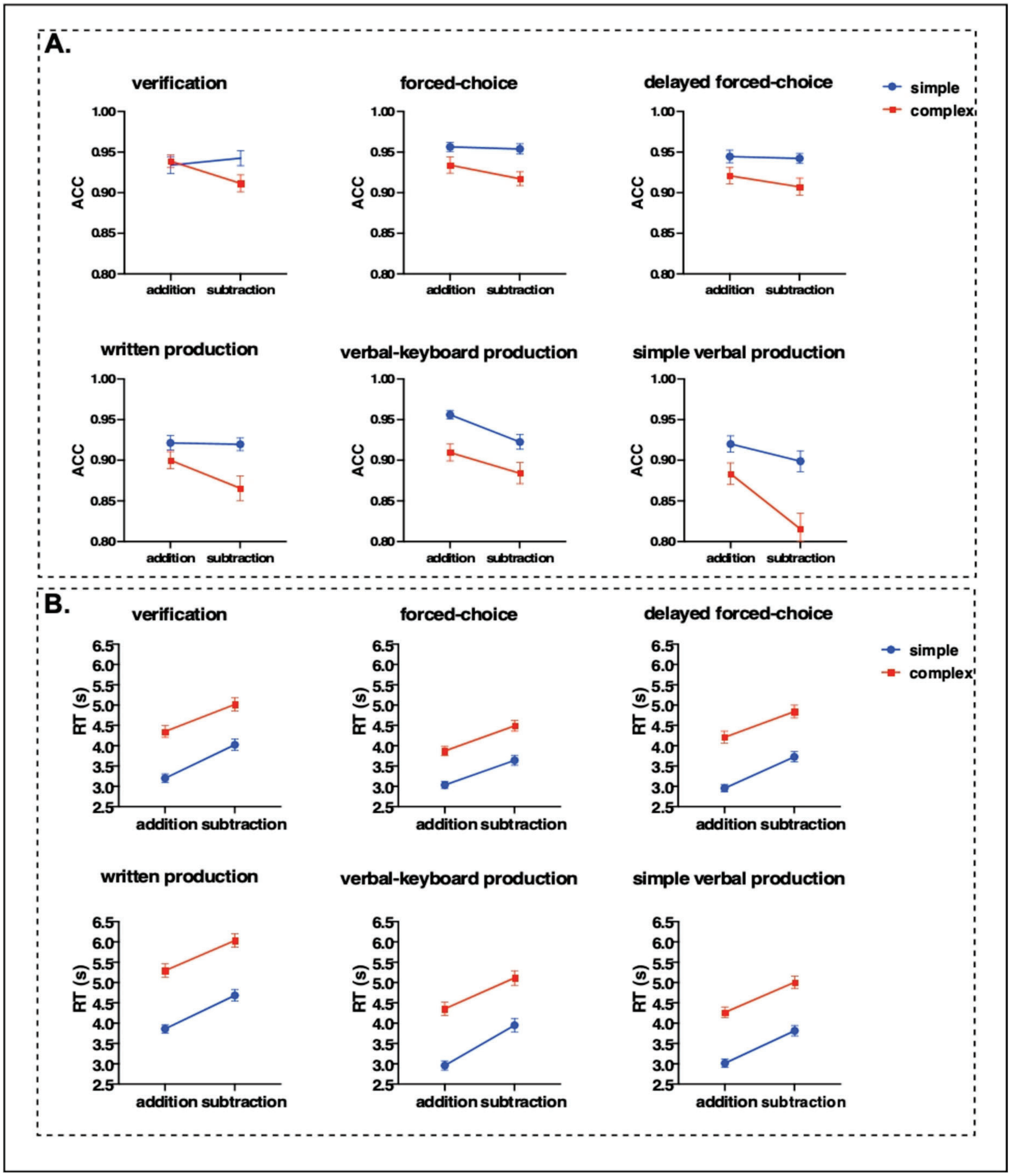


Fig. 3. Arithmetic effects for operation and difficulty across paradigms.
Notes. The upper panel is raw accuracy distribution, and the lower panel is reaction time distribution across six paradigms. Error bars represent the standard error of the mean.

this consistency across paradigms highlights the robustness of the difficulty effects in arithmetic. As an exception, the verification paradigm did not reveal a significant difficulty effect in accuracy (possibly due to ceiling effects).

Interaction effects. In none of the paradigms, paradigm-specific analyses did reveal a significant interaction between operation and difficulty or sufficient evidence to support an interaction (for details see Supplementary Materials, Table S5).

Distractor effects. For the forced-choice paradigm, the paradigm-specific analysis of both accuracy and reaction time showed significant distractor effects (see Tables 3 & 4; Supplementary Materials, Fig. S2), indicating that calculation was faster and more accurate when the distractor differed from the target in the unit (±2) than in the decade (±10) (e.g., Artemenko, Pixner, Moeller, & Nuerk, 2018; Moeller et al., 2011a). These results indicate that distractors influence the solution process, affecting both the speed and accuracy of responses in forced-choice paradigm.

For the delayed forced-choice paradigm, moderate evidence against a distractor effect was found for the first reaction time (calculation phase), whereas significant distractor effects were observed for the second reaction time (response phase) and even for accuracy (see Table 4; Supplementary Materials, Fig. S2). The delayed forced-choice

Table 5
Exploratory results of the second reaction time in the delayed forced-choice paradigm.

factor	ANOVA			Bayesian ANOVA			
	F	p	η_p^2	P_{incl}	$P_{incl\|data}$	BF_{incl}	BF_{excl}
operation	2.34	0.131	0.04	0.26	0.02	0.43	2.33
difficulty	11.19	< **0.001**	0.15	0.26	0.06	**18.90**	0.05
distractor	12.26	< **0.001**	0.15	0.26	0.51	**15.65**	0.06
operation × difficulty	16.23	< **0.001**	0.21	0.26	0.92	**31.67**	0.03
operation × distractor	3.50	0.066	0.05	0.26	0.36	0.65	1.55
difficulty × distractor	0.08	0.776	0.00	0.26	0.15	0.18	**5.57**
operation × difficulty × distractor	0.04	0.849	0.00	0.05	0.01	0.18	**5.56**

Notes. P_{incl} represents the prior inclusion probability of an effect, $P_{incl|data}$ represents the inclusion probability given the data. BF_{incl} stands for the Bayes Factor for inclusion, indicating evidence for a difference; BF_{excl} stands for the Bayes factor for exclusion, indicating evidence for a null effect.

paradigm was designed to separate mental calculation (first reaction time) from decision-making (second reaction time), as suggested by previous research (Zbrodoff & Logan, 1990). Our findings demonstrate that distractor effects are specific to the response phase and, by definition, do not influence the calculation phase. While the calculation phase may resemble production tasks to a certain extent, the response phase involves additional decision-related processes that are influenced by distractor information. Importantly, these results align with the expectation that different cognitive processes may emerge in distinct phases of the task and these phases can influence each other.

3.3. Magnitudes of arithmetic effects (effect size)

While all six paradigms showed similar medium effect sizes for the operation effect and large effect sizes for the difficulty effect in reaction time, effect sizes for accuracy were higher in the two verbal production paradigms (see Table 2, Fig. 4; Supplementary Materials, Table S6). This means that differences in calculation speed are rather independent of the paradigm, while for precision of arithmetic matters which paradigm is chosen. In production, the open response format without any given answers is more sensitive to errors during calculation, while in decision paradigms, multiple strategies can be employed, leading to larger variance. These findings emphasize the importance of considering both the type of the paradigm and the dependent variable being measured when interpreting results in arithmetic research. The consistency in reaction time of the investigated arithmetic effects across paradigms supports its stability as a measure of arithmetic processing, whereas the variability in accuracy of the investigated arithmetic effects highlights the need for careful selection of paradigms, especially when the goal is to assess the precision of arithmetic performance.

4. Discussion

The present study provides critical insights into the similarities and differences in arithmetic effects and performance concerning six paradigms. Our results showed that (1) the arithmetic performance level differed between paradigms (consistently higher accuracy in decision than production paradigms; inconsistent pattern for reaction time); (2) operation and difficulty effects are replicated in all six paradigms in terms of reaction time but not for accuracy; (3) effect sizes for accuracy were paradigm-dependent (higher effect sizes for verbal production paradigms), while effect sizes for reaction time were not.

Interestingly, our study revealed both consistencies and inconsistencies across paradigms, contingent on several factors: the specific effect of interest, the theoretical question addressing only the existence of an effect or also its size, and the choice of the dependent variable (reaction time vs. accuracy). Consequently, it is not always appropriate to simply generalize from one paradigm to cognitive and numerical processes in general, but in many cases – such as when the study aims to just demonstrate the existence of an effect – it might be sufficiently consistent with other paradigms. By using these

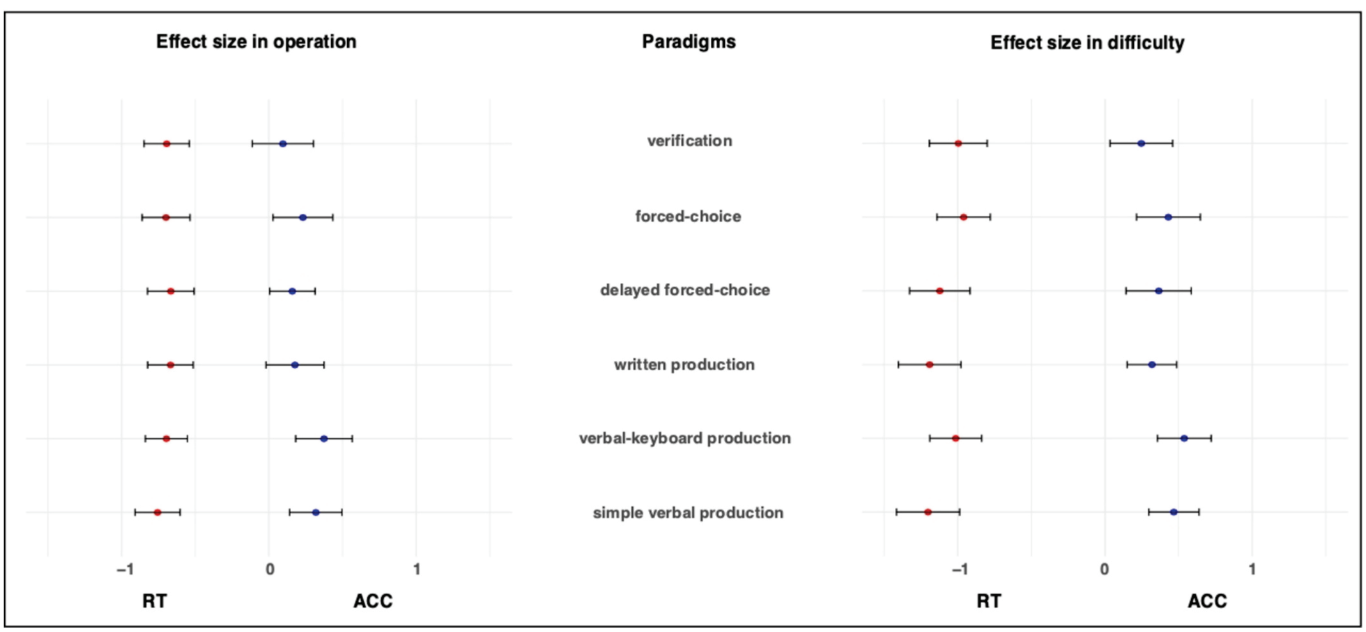


Fig. 4. Effect sizes for operation and difficulty effects across paradigms.
Notes. The left panel shows effect sizes (Cohen's *d*) for operation in terms of reaction time (RT; red dots) and accuracy (ACC; blue dots) with 90 % CI. The right panel shows effect sizes for difficulty in terms of reaction time (RT; red dots) and accuracy (ACC; blue dots) with 90 % CI. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

differentiations, the results are discussed in terms of their theoretical as well as methodological significance and implications.

4.1. Cognitive processes of different paradigms

Paradigm-specificity is especially important when different paradigms engage distinct cognitive processes. In the current study, accuracy was consistently higher in decision paradigms compared to production paradigms. While this is partly due to simple guessing, because the chance is higher in decision paradigms, this is likely not the full story. The the observed distractor effects highlight how the presence of given answers (even if incorrect) shapes arithmetic processing in decision paradigms. Responses can be based on rejecting the wrong answer, a pattern that inherently can not be observed in production paradigms.

In the verification paradigm, the decision-making process may rely more on plausibility judgments than pure arithmetic calculation (Campbell & Tarling, 1996; Duverne & Lemaire, 2005; Hinault & Lemaire, 2016). In consequence, the verification paradigm does not exclusively measure pure calculation processes (like production paradigms), but rather a combination of estimation, retrieval, and evaluation. Our findings further underscore this assumption. Specifically, the verification paradigm yielded lower sensitivity to the difficulty effect in accuracy observed in all other paradigms. Moreover, reaction times in verification were slower than in other decision paradigms, where participants actively reject distractors. These results suggest that cognitive processes involved in verification are distinct from those used in real-world arithmetic, emphasizing the need for caution when generalizing findings from this paradigm to broader assessments of arithmetic ability.

In the forced-choice paradigm, two potential answers are given and the correct one should be selected, leading to distractor effects (e.g., Artemenko, Pixner, Moeller, & Nuerk, 2018; Moeller et al., 2011a). The presence of a distractor creates a scenario where multiple strategies can be employed: either verifying the correct answer by mental calculation or rejecting the distractor by recognizing that it is incorrect (Moeller et al., 2011a). This flexibility in strategy use likely contributes to the high speed and accuracy observed in forced-choice paradigms compared to other paradigms, as the most efficient approach (e.g., short-cut strategies) can be employed depending on the complexity of the problem. Therefore, our results highlight that the given answers in a forced-choice paradigm facilitate arithmetic processing due to an adaptive strategy use.

Distractor effects were also found in the delayed forced-choice paradigm. This can be explained by the unique structure of the paradigm, where calculation and decision phases might influence each other. During the initial calculation phase, individuals may generate a preliminary, unverified result with a certain level of uncertainty, anticipating the presentation of both a correct answer and a distractor in the response phase. The cognitive load during this phase is likely reduced because participants know they will have the opportunity to verify or adjust their initial answer once the options are presented. Our results suggest that this anticipation allows to strategically offload some of the mental effort, treating the calculation phase as an intermediate step rather than requiring complete production of the solution. This process is similar to lexical ambiguity resolution in language comprehension, where contextual cues help disambiguate meanings (Gorfein, 2001). Readers generate preliminary interpretations of ambiguous words and select the most appropriate interpretation based on the overall meaning of the sentence or discourse (Bousquet et al., 2020; Swaab et al., 2003; Zempleni et al., 2007). Similarly, in the delayed forced-choice paradigm an approximate answer might be generated during calculation, knowing that the task context allows for refinement during the response phase. Consequently, the processes involved in the delayed forced-choice paradigm diverge from those in production paradigms.

This distinction between decision and production paradigms has significant implications for our understanding of mental arithmetic. In production paradigms, we predominantly measure pure calculation processes, as no answers are given. Conversely, in decision paradigms, calculation processes can be adjusted based on the provided answer options due to the complex interplay of cognitive processes besides calculation (estimation, short-cut strategies and evaluative processes). This explains why decision paradigms, like forced-choice, tend to reveal better performance compared to production paradigms, where such external aids are absent and individuals must rely solely on their calculation abilities.

4.2. Existence and magnitude of arithmetic effects across paradigms

Whether or not an arithmetic effect can be detected as well as the size of the effect was consistent for all paradigms when concerning reaction times – but not accuracy. For reaction time, we observed operation and difficulty effects in two-digit arithmetic across all six paradigms, which were also comparable in their effect size. In this regard, effect-based approaches in arithmetic research can be treated as paradigm-independent. The operation effect, with subtraction problems taking longer to solve than addition problems, was replicated in all paradigms (Campbell et al., 2006; De Smedt et al., 2009; Klein et al., 2014). Moreover, the carry and borrow effects were replicated indicating higher difficulty for arithmetic involving carrying or borrowing compared to simpler arithmetic without carrying and borrowing (Artemenko, Soltanlou, Dresler, et al., 2018; Deschuyteneer et al., 2005; Imbo & LeFevre, 2010). The replication of the arithmetic effects regardless of the paradigm shows the robustness of these effects and the similarity of the paradigms in additional time for processing task difficulty.

However, when it comes to accuracy, the choice of paradigm plays a critical role. Our findings indicate that the verbal production paradigms yielded higher effect sizes for the arithmetic effects under investigation and the operation effect could only be detected in these paradigms. The open response format of production paradigms was more sensitive to errors during calculation, as compared to decision paradigms where possible answers are given. This sensitivity is consistent with the Problem-Solving Theory (Ashcraft, 1992), which suggests that generating an answer (as in production paradigms) requires deeper processing and more cognitive effort compared to selecting an answer (as in decision paradigms). Therefore, differences in accuracy can be better detected in production than decision paradigms.

4.3. Limitations

The generalizability of the current findings is constrained by factors related to language, age, and task design. First, language significantly influences two-digit number processing and arithmetic performance. Our sample speaks German, which is characterized by number word inversion for two-digit numbers, where the unit is spoken before the decade (e.g., "einundzwanzig" literally translates to "one-and-twenty"). In languages with number word inversion, the carry effect is larger (Lewis et al., 2020). Future research should address the universality or language-specificity of these phenomena by investigating these effects in languages without inversion or in those with a more transparent verbal place-value structure.

Second, strategy use might be dependent on development. For instance, school children are particularly sensitive to the testing format—whether it is multiple choice or open answer (Danili & Reid, 2006; Frederiksen, 1984). Therefore, future research should include a broader age range (e.g., Avcil & Artemenko, 2023) to evaluate whether the observed behavioral patterns generalize to other developmental stages, such as children or the elderly.

Third, further research is needed to determine if the observed paradigm-(in)dependent behavioral patterns are specific to the arithmetic effects studied or generalize to other arithmetic operations, levels of complexity, and mathematical tasks. For instance, Dewi et al. (2021) studied the solutions of algebraic equations and did not find differences between paradigms. Therefore, the paradigm-(in-)consistency observed here for arithmetic processing might or might not generalize to all domains of mathematical cognition.

Finally, our findings suggest that arithmetic effects are partially similar (reaction time) and partially different (accuracy) between paradigms. What is more, arithmetic effects might even vary within a paradigm. In the Ironman paradigm by Roth et al. (2024), group-level numerical effects were replicated, but the effects were reliably present only in a subset of participants, and varied even from day to day at an individual level. In our study, we also observed effects at the group level, but the stability of these effects at the individual level remains a question of future research. Additionally, the cognitive processes engaged may fluctuate based on task demands, leading to performance variability even within a single paradigm (cf. Roth et al. (2024). In line with the Dual-Process Theory (Evans & Stanovich, 2013), individuals might alternate between fast, intuitive processing (system 1) and slower, more deliberate processing (system 2), depending on their familiarity with the task or the complexity of the problem. This might have also influenced our experimental design, as the same arithmetic task was repeated several times in different paradigms. Thus, within- and between-paradigm variabilities need to be recognized and accounted for in future research.

5. Conclusions

The key finding of this study is that some effects in arithmetic are relatively consistent across paradigms, while others are more paradigm-specific. Whether this is the case depends on the effect of interest, the dependent variable (accuracy vs. reaction time), and the importance of the effect's existence versus its size. For instance, the operation and difficulty effects are robust across all paradigms in terms of reaction time. Conversely, accuracy depends on the type of paradigm used: verbal production paradigms show higher effect sizes and are more prone to errors than decision paradigms with given answers that allow for more strategic approaches.

This has implications for paradigm selection in cognitive research, as the choice of paradigm can in some cases significantly impact the nature and interpretation of the cognitive processes being studied. In conclusion, our research underscores the notion that "arithmetic is not arithmetic" when assessed across different paradigms, emphasizing that the choice of paradigm can shape both the observed outcomes and the theoretical conclusions drawn about arithmetic processing.

CRediT authorship contribution statement

Xinru Yao: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Christina Artemenko:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Conceptualization. **Yunfeng He:** Writing – review & editing, Methodology, Conceptualization. **Hans-Christoph Nuerk:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no conflicts of interest.

Acknowledgments

We would like to thank the participants in this study. We are also grateful to Annalena Wels for her help in data collection and Sebastian Sandbrink for proofreading the manuscript.

XY, CA, and HCN are members of the LEAD Graduate School & Research Network (GSC1028, funded by the Excellence Initiative of the German federal and state governments). CA was supported by the European Social Fund and the Ministry of Science, Research and the Arts Baden-Wuerttemberg, by the German Research Foundation (DFG, grant number: 468460838, AR 1500/1-1; grant number: 513458453, AR 1500/2-1), and by the Tuebingen Postdoc Academy for Research on Education (PACE) at the Hector Research Institute of Education Sciences and Psychology. YH was supported by the Ministry of Education Humanities and Social Science Project (18YJC190006) in China.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.106060.

Data availability

all data and materials have been uploaded on OSF

References

Amalric, M., & Cantlon, J. F. (2022). Common neural functions during Children's learning from naturalistic and controlled mathematics paradigms. *Journal of Cognitive Neuroscience, 34*(7), 1164–1182.

Artemenko, C. (2018). *Neurocognitive foundations of arithmetic complexity in adults and children*. Eberhard Karls Universität Tübingen.

Artemenko, C., Moeller, K., Huber, S., & Klein, E. (2015). Differential influences of unilateral tDCS over the intraparietal cortex on numerical cognition. *Frontiers in Human Neuroscience, 9*, 110.

Artemenko, C., Pixner, S., Moeller, K., & Nuerk, H. C. (2018). Longitudinal development of subtraction performance in elementary school. *British Journal of Developmental Psychology, 36*(2), 188–205.

Artemenko, C., Soltanlou, M., Dresler, T., Ehlis, A.-C., & Nuerk, H.-C. (2018). The neural correlates of arithmetic difficulty depend on mathematical ability: Evidence from combined fNIRS and ERP. *Brain Structure and Function, 223*(6), 2561–2574. https://doi.org/10.1007/s00429-018-1618-0

Artemenko, C., Soltanlou, M., Ehlis, A.-C., Nuerk, H.-C., & Dresler, T. (2018). The neural correlates of mental arithmetic in adolescents: A longitudinal fNIRS study. *Behavioral and Brain Functions, 14*(1), 1–13.

Ashcraft, M. (1984). The production and verification tasks in mental addition: An empirical comparison*1. *Developmental Review, 4*(2), 157–170. https://doi.org/10.1016/0273-2297(84)90005-4

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1), 75–106. https://doi.org/10.1016/0010-0277(92)90051-I

Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4*(5), 527.

Avcil, M., & Artemenko, C. (2023). Development of arithmetic across the lifespan: A registered report. https://osf.io/preprints/psyarxiv/rkf6y.

Bayot, M., Dujardin, K., Tard, C., Defebvre, L., Bonnet, C. T., Allart, E., & Delval, A. (2018). The interaction between cognition and motor control: A theoretical framework for dual-task interference effects on posture, gait initiation, gait and turning. *Neurophysiologie Clinique, 48*(6), 361–375. https://doi.org/10.1016/j.neucli.2018.10.003

Bousquet, K., Swaab, T. Y., & Long, D. L. (2020). The use of context in resolving syntactic ambiguity: Structural and semantic influences. *Language, Cognition and Neuroscience, 35*(1), 43–57. https://doi.org/10.1080/23273798.2019.1622750

Brysbaert, M., Fias, W., & Noël, M.-P. (1998). The Whorfian hypothesis and numerical cognition: Istwenty-four'processed in the same way asfour-and-twenty'? *Cognition, 66*(1), 51–77.

Campbell, J. I. (1987). Production, verification, and priming of multiplication facts. *Memory & Cognition, 15*(4), 349–364.

Campbell, J. I., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition, 34*(3), 633–647.

Campbell, J. I., & Fugelsang, J. (2001). Strategy choice for arithmetic verification: Effects of numerical surface form. *Cognition, 80*(3), B21–B30.

Campbell, J. I., & Tarling, D. P. (1996). Retrieval processes in arithmetic production and verification. *Memory & Cognition, 24*(2), 156–172.

Cantlon, J. F. (2020). The balance of rigor and reality in developmental neuroscience. *Neuroimage, 216*, Article 116464.

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 747. https://doi.org/10.1037/0278-7393.33.4.747

Dackermann, T., Kroemer, L., Nuerk, H.-C., Moeller, K., & Huber, S. (2018). Influences of presentation format and task instruction on children's number line estimation. *Cognitive Development, 47*, 53–62. https://doi.org/10.1016/j.cogdev.2018.03.001

Danili, E., & Reid, N. (2006). Cognitive factors that can potentially affect pupils' test performance. *Chemistry Education Research and Practice, 7*(2), 64–83.

De Smedt, B., Grabner, R. H., & Studer, B. (2009). Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. *Experimental Brain Research, 195*(4), 635–642. https://doi.org/10.1007/s00221-009-1839-9

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.

Deschuyteneer, M., De Rammelaere, S., & Fias, W. (2005). The addition of two-digit numbers: Exploring carry versus no-carry problems. *Psychology Science, 47*(1), 74–83.

Dewi, J. D., Bagnoud, J., & Thevenot, C. (2021). Do production and verification tasks in arithmetic rely on the same cognitive mechanisms? A test using alphabet arithmetic. *The Quarterly Journal of Experimental Psychology, 74*(12), 2182–2192. https://doi.org/10.1177/17470218211022635

Duverne, S., & Lemaire, P. (2005). *Aging and mental arithmetic*. Psychology Press.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition, 6*(2), 231–259.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*(3), 193.

Gorfein, D. S. (2001). On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. *American Psychological Association.* https://doi.org/10.1037/10459-000

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & 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. https://doi.org/10.1016/j.neuropsychologia.2008.10.013

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition, 30*, 447–455.

Hinault, T., & Lemaire, P. (2016). What does EEG tell us about arithmetic strategies? A review. *International Journal of Psychophysiology, 106*, 115–126. https://doi.org/10.1016/j.ijpsycho.2016.05.006

Imbo, I., & LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese-and Canadian-educated adults. *Memory & Cognition, 38*(2), 176–185. https://doi.org/10.3758/MC.38.2.176

Imbo, I., & Vandierendonck, A. (2007). The role of phonological and executive working memory resources in simple arithmetic strategies. *European Journal of Cognitive Psychology, 19*(6), 910–933. https://doi.org/10.1080/09541440601051571

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological Research, 71*(4), 467–483. https://doi.org/10.1007/s00426-006-0044-8

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32*, 662–673.

Jeffreys, H. (1998). *The theory of probability*. Oxford University Press.

Klein, E., Huber, S., Nuerk, H.-C., & Moeller, K. (2014). Operational momentum affects eye fixation behaviour. *Quarterly Journal of Experimental Psychology, 67*(8), 1614–1625.

Klein, E., Nuerk, H.-C., Wood, G., Knops, A., & Willmes, K. (2009). The exact vs. approximate distinction in numerical cognition may not be exact, but only approximate: How different processes work together in multi-digit addition. *Brain and Cognition, 69*(2), 369–381. https://doi.org/10.1016/j.bandc.2008.08.031

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

LeFevre, J. A., Shanahan, T., & DeStefano, D. (2004). The tie effect in simple arithmetic: An access-based account. *Memory & Cognition, 32*(6), 1019–1031. https://doi.org/10.3758/bf03196878

Lewis, C. A., Bahnmueller, J., Wesierska, M., Moeller, K., & Göbel, S. M. (2020). Inversion effects on mental arithmetic in English-and polish-speaking adults. *Quarterly Journal of Experimental Psychology, 73*(1), 91–103.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Moeller, K., Klein, E., & Nuerk, H. C. (2011b). Three processes underlying the carry effect in addition–evidence from eye tracking. *British Journal of Psychology, 102*(3), 623–645. https://doi.org/10.1111/j.2044-8295.2011.02034.x

Moeller, K., Klein, E., & Nuerk, H.-C. (2011a). (no) small adults: Children's processing of carry addition problems. *Developmental Neuropsychology, 36*(6), 702–720.

Nuerk, H.-C., Geppert, B. E., van Herten, M., & Willmes, K. (2002). On the impact of different number representations in the number bisection task. *Cortex, 38*(5), 691–715. https://doi.org/10.1016/s0010-9452(08)70038-8

Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). *Multi-digit number processing: Overview, conceptual clarifications, and language influences*. Oxford University Press.

Pletzer, B., Scheuringer, A., & Harris, T. (2016). Spacing and presentation modes affect the unit-decade compatibility effect during number comparison. *Experimental Psycholog, 63*(3), 189–195. https://doi.org/10.1027/1618-3169/a000326

Roth, L., Jordan, V., Schwarz, S., Willmes, K., Nuerk, H.-C., van Dijck, J.-P., & Cipora, K. (2024). Don't SNARC me now! Intraindividual variability of cognitive phenomena – Insights from the ironman paradigm. *Cognition, 248*, Article 105781. https://doi.org/10.1016/j.cognition.2024.105781

Segal, A. (2011). *Do gestural interfaces promote thinking? Embodied interaction: Congruent gestures and direct touch promote performance in math*. Columbia University.

Seyler, D. J., Kirk, E. P., & Ashcraft, M. H. (2003). Elementary subtraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1339.

Siegler, R. S. (2024). Connecting learning environments to learning: Two examples from children's mathematics. *Developmental Review*. , Article 101138. https://doi.org/10.1016/j.dr.2024.101138

Siegler, R. S., & Oppenzato, C. O. (2021). Missing input: How imbalanced distributions of textbook problems affect mathematics learning. *Child Development Perspectives, 15*(2), 76–82. https://doi.org/10.1111/cdep.12402

Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language, 86*(2), 326–343. https://doi.org/10.1016/S0093-934X(02)00547-3

Thevenot, C., Dewi, J. D. M., Bagnoud, J., Uittenhove, K., & Castel, C. (2020). Scrutinizing patterns of solution times in alphabet-arithmetic tasks favors counting over retrieval models. *Cognition, 200*, Article 104272. https://doi.org/10.1016/j.cognition.2020.104272

Trbovich, P. L., & LeFevre, J.-A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31*(5), 738–745.

Tschentscher, N., & Hauk, O. (2014). How are things adding up? Neural differences between arithmetic operations are due to general problem solving strategies. *Neuroimage, 92*, 369–380.

Uittenhove, K., & Lemaire, P. (2014). Numerical cognition during cognitive aging. In A. D. Roi Cohen Kadosh (Ed.), *The Oxford handbook of numerical cognition* (pp. 345–364). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.045.

Zbrodoff, N. J. (1999). Effects of counting in alphabet arithmetic: Opportunistic stopping and priming of intermediate steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 299.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General, 115*(2), 118.

Zbrodoff, N. J., & Logan, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 83.

Zempleni, M.-Z., Renken, R., Hoeks, J. C., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *Neuroimage, 34*(3), 1270–1279. https://doi.org/10.1016/j.neuroimage.2006.09.048

A1.2 Supplementary Materials for Study 1

These supplementary materials belong to the following publication:

Yao, X., Artemenko, C., He, Y., & Nuerk, H.-C. (2025). Arithmetic is not arithmetic: Paradigm matters for arithmetic effects. Cognition, 256, 106060. https://doi.org/10.1016/j.cognition.2024.106060.

Supplementary Material

Title: Arithmetic is not arithmetic: Paradigm matters for arithmetic effects

Authors: Xinru Yao, Christina Artemenko, Yunfeng He, Hans-Christoph Nuerk

Abbreviations: ACC – accuracy, RT – reaction time

Table S1. Number of excluded participants and trials per paradigm

	verification	**forced-choice**	**delayed forced-choice**	**written production**	**Verbal-keyboard production**	**Simple verbal production**
Number of excluded participants per paradigm						
(1) Drop out	1	1	1	1	1	1
(2) Missing data	0	0	0	1	0	1
(3) ACC	2 (< 75%)	1 (< 75%)	1 (< 75%)	0 (< 50%)	0 (< 50%)	0 (< 50%)
(4) RT outlier	4	5	1	2	0	1
total	6	6	2	2	0	1
Percentage of excluded trials per paradigm						
(1) incorrect trials	4%	7%	8%	10%	8%	12%
(2) RT < 200 ms	0%	0%	0%	0%	0%	0%
(3) RT outlier	1%	3%	10%	10%	5%	4%
(4) RT duration outlier	-	-	0%	0%	0%	-

Notes. According to the exclusion criteria of participants, in decision paradigms (verification, forced-choice, delayed forced-choice), participants with an accuracy below 75% were case-wise excluded; in production paradigms (written production, verbal-keyboard production, simple verbal production), participants with an accuracy below 50% were case-wise excluded. Data were excluded on a trial level for RT analysis. Outliers were detected by *Median* ± 3 *MAD*. In delayed forced-choice, written production, and verbal-keyboard paradigms, response duration was determined as the duration between two button presses. Note that percentages are rounded to integer numbers.

Table S2. Descriptive results for arithmetic performance in six paradigms

		verification *N* **= 59**	**forced-choice** *N* **= 59**	**delayed forced-choice** *N* **= 63**	**written production** *N* **= 62**	**verbal-keyboard production** *N* **= 65**	**simple verbal production** *N* **= 63**
ACC	Overall	0.93 (*0.04*)	0.94 (*0.04*)	0.93 (*0.05*)	0.90 (*0.06*)	0.92 (*0.05*)	0.88 (*0.09*)
	Addition simple	0.93 (*0.08*)	0.96 (*0.04*)	0.95 (*0.06*)	0.92 (*0.07*)	0.96 (*0.04*)	0.92 (*0.08*)
	Addition complex	0.94 (*0.06*)	0.93 (*0.08*)	0.92 (*0.08*)	0.90 (*0.08*)	0.91 (*0.08*)	0.88 (*0.11*)
	Subtraction simple	0.94 (*0.07*)	0.95 (*0.05*)	0.94 (*0.05*)	0.92 (*0.06*)	0.92 (*0.07*)	0.90 (*0.10*)
	Subtraction complex	0.91 (*0.08*)	0.96 (*0.04*)	0.91 (*0.08*)	0.87 (*0.12*)	0.88 (*0.10*)	0.82 (*0.15*)
Logit ACC	Overall	2.65 (*0.46*)	2.64 (*0.50*)	2.52 (*0.53*)	2.23 (*0.57*)	2.42 (*0.55*)	2.14 (*0.73*)
	Addition simple	2.72 (*0.98*)	2.82 (*0.70*)	2.72 (*0.81*)	2.38 (*0.76*)	2.80 (*0.69*)	2.43 (*0.84*)
	Addition complex	2.66 (*0.86*)	2.64 (*0.90*)	2.44 (*0.84*)	2.23 (*0.86*)	2.33 (*0.87*)	2.14 (*0.95*)
	Subtraction simple	2.81 (*0.95*)	2.80 (*0.71*)	2.60 (*0.70*)	2.34 (*0.71*)	2.47 (*0.85*)	2.28 (*0.96*)
	Subtraction complex	2.40 (*0.94*)	2.31 (*0.71*)	2.30 (*0.85*)	1.98 (*0.95*)	2.06 (*0.82*)	1.68 (*1.08*)
RT (s)	Overall	4.15 (0.*97*)	3.76 (0.*79*)	3.93 (0.*95*)	4.97 (*1.05*)	4.10 (*1.17*)	4.03 (0.*92*)
	Addition simple	3201 (0.*84*)	3036 (0.*64*)	2956 (0.*75*)	3862 (0.*81*)	2959 (0.*91*)	3016 (0.*80*)
	Addition complex	4.35 (*1.09*)	3.87 (0.*86*)	4.21 (*1.17*)	5.30 (*1.32*)	4.35 (*1.30*)	4.27 (*1.04*)
	Subtraction simple	4.03 (*1.09*)	3.64 (0.*96*)	3.73 (*1.01*)	4.69 (*1.15*)	3.95 (*1.34*)	3.82 (*1.00*)
	Subtraction complex	5.02 (*1.23*)	4.49 (*1.02*)	4.84 (*1.23*)	6.04 (*1.33*)	5.11 (*1.43*)	5.01 (*1.19*)

Notes. The upper part shows mean accuracy with standard deviation, the middle part shows logit-transformed mean accuracy with standard deviation, and the lower part shows mean reaction time (with second as unit) with standard deviation.

Table S3a. Pairwise comparisons of accuracy for paradigm (three-way ANOVA and Bayesian ANOVA)

paradigms		ΔM	t	P_{holm}	Cohen's d [95% CI]		BF_{10}	BF_{01}
verification	forced-choice	-0.04	-0.57	.823	-0.046	[-0.327, 0.235]	0.09	**11.51**
	delayed forced-choice	0.05	0.83	.823	0.061	[-0.220, 0.342]	0.10	**10.43**
	written production	0.33	5.42	< **.001**	0.396	[0.093, 0.699]	**> 100**	0.00
	verbal-keyboard production	0.15	1.86	.413	0.173	[-0.112, 0.459]	0.37	0.74
	simple verbal production	0.43	4.78	< **.001**	0.519	[0.201, 0.836]	**> 100**	0.00
forced-choice	delayed forced-choice	0.09	1.42	.644	0.107	[-0.175, 0.390]	0.20	**4.90**
	written production	0.37	5.22	< **.001**	0.442	[0.134, 0.750]	**> 100**	0.00
	verbal-keyboard production	0.18	2.49	.128	0.220	[-0.068, 0.508]	**4.28**	0.23
	simple verbal production	0.47	4.41	< **.001**	0.565	[0.241, 0.889]	**> 100**	0.00
delayed forced-choice	written production	0.28	4.29	**.001**	0.335	[0.038, 0.632]	**> 100**	0.00
	verbal-keyboard production	0.09	1.57	.616	0.112	[-0.170, 0.395]	0.25	**4.00**
	simple verbal production	0.38	3.91	**.003**	0.458	[0.148, 0.768]	**> 100**	0.00
written production	verbal-keyboard production	-0.19	-2.48	.128	-0.222	[-0.510, 0.066]	**7.84**	0.13
	simple verbal production	0.10	1.22	.685	0.123	[-0.160, 0.406]	0.25	**4.05**
verbal-keyboard production	simple verbal production	0.29	2.79	.065	0.345	[0.048, 0.643]	**> 100**	0.01

Notes. Bonferroni-Holm correction was applied to the pairwise comparisons. BF_{10} stands for the Bayes Factor comparing the alternative hypothesis (H1) to the null hypothesis (H0), indicating evidence for a difference. BF_{01} stands for the Bayes Factor comparing the null hypothesis (H0) to the alternative hypothesis (H1), indicating evidence for a null effect.

Table S3b. Pairwise comparisons of reaction time for paradigm (three-way ANOVA and Bayesian ANOVA)

paradigms		ΔM	t	$p_{corrected}$	Cohen's d	[95% CI]	BF_{10}	BF_{01}
verification	forced-choice	337.43	4.44	**< .001**	0.337	[0.051, 0.623]	**> 100**	0.00
	delayed forced-choice	221.72	2.36	.177	0.222	[-0.055, 0.498]	**57.92**	0.02
	written production	-830.99	-7.18	**< .001**	-0.830	[-1.192, -0.469]	**> 100**	0.00
	verbal-keyboard production	166.78	1.54	.784	0.167	[-0.107, 0.440]	1.60	0.63
	simple verbal production	106.19	1.11	.821	0.106	[-0.165, 0.377]	0.34	2.91
forced-choice	delayed forced-choice	-115.71	-1.51	.784	-0.116	[-0.387, 0.155]	1.16	0.87
	written production	-1168.42	-14.55	**< .001**	-1.168	[-1.600, -0.735]	**> 100**	0.00
	verbal-keyboard production	-170.65	-2.02	.342	-0.171	[-0.444, 0.103]	**9.04**	0.11
	simple verbal production	-231.24	-2.87	.052	-0.231	[-0.508, 0.046]	**> 100**	0.00
delayed forced-choice	written production	-1052.71	-13.91	**< .001**	-1.052	[-1.459, -0.645]	**> 100**	0.00
	verbal-keyboard production	-54.94	-0.613	.994	-0.055	[-0.324, 0.214]	0.13	**7.77**
	simple verbal production	-115.53	-1.39	.784	-0.115	[-0.386, 0.156]	1.06	0.94
written production	verbal-keyboard production	997.77	12.47	**< .001**	0.997	[0.602, 1.392]	**> 100**	0.00
	simple verbal production	937.18	9.64	**< .001**	0.937	[0.554, 1.319]	**> 100**	0.00
verbal-keyboard production	simple verbal production	-60.59	-0.68	.994	-0.061	[-0.330, 0.209]	0.15	**6.87**

Notes. Bonferroni-Holm correction was applied to the pairwise comparisons. BF_{10} stands for the Bayes Factor comparing the alternative hypothesis (H1) to the null hypothesis (H0), indicating evidence for a difference. BF_{01} stands for the Bayes Factor comparing the null hypothesis (H0) to the alternative hypothesis (H1).

Table S4a. Pairwise comparisons of accuracy for operation and difficulty (two-way ANOVAs and Bayesian ANOVAs)

factor	paradigm	ΔM	t	p_{holm}	Cohen's d [95% CI]		BF_{10}	BF_{01}
operation	verification	0.09	0.76	.448	0.095	[-0.153, 0.344]	0.13	**7.62**
	forced-choice	0.18	1.92	.060	0.231	[-0.012, 0.474]	0.94	1.07
	delayed forced-choice	0.13	1.73	.088	0.159	[-0.026, 0.344]	0.36	2.78
	written production	0.15	1.51	.136	0.177	[-0.057, 0.412]	0.35	2.83
	verbal-keyboard production	0.30	3.38	**.001**	0.374	[0.145, 0.604]	**26.11**	0.04
	simple verbal production	0.31	3.10	**.003**	0.319	[0.107, 0.530]	**18.24**	0.06
difficulty	verification	0.23	1.96	.055	0.246	[-0.007, 0.500]	0.53	1.91
	forced-choice	0.33	3.47	**< .001**	0.430	[0.172, 0.689]	**> 100**	0.01
	delayed forced-choice	0.29	2.81	**.007**	0.365	[0.100, 0.630]	**9.82**	0.10
	written production	0.26	3.30	**.002**	0.319	[0.119, 0.519]	**12.61**	0.08
	verbal-keyboard production	0.44	5.43	**< .001**	0.538	[0.320, 0.756]	**> 100**	0.00
	simple verbal production	0.45	4.98	**< .001**	0.467	[0.263, 0.672]	**> 100**	0.00
distractor	forced-choice	0.66	7.83	**< .001**	0.714	[0.490, 0.937]	**> 100**	0.00
	delayed forced-choice	0.31	3.58	**< .001**	0.311	[0.130, 0.492]	**70.41**	0.00

Notes. ΔM was logit-transformed accuracy in addition minus logit-transformed accuracy in subtraction (for operation), logit-transformed accuracy in simple minus logit-transformed accuracy in complex (for difficulty), and logit-transformed accuracy in distractor distance 2 minus logit-transformed accuracy in distractor distance 10 (for distractor). Bonferroni-Holm correction was applied to the pairwise comparisons. BF_{10} stands for the Bayes Factor comparing the alternative hypothesis (H_1) to the null hypothesis (H_0), BF_{01} stands for the Bayes Factor comparing the null hypothesis (H_0) to the alternative hypothesis (H_1).

Table S4b. Pairwise comparisons of reaction time for operation, difficulty, and distractor (frequentist ANOVAs and Bayesian ANOVAs)

factor	paradigm	ΔM	t	p_{holm}	Cohen's d	[95% CI]	BF_{10}
operation	verification	-745.68	-10.45	**< .001**	-0.694	[-0.879, -0.509]	**> 100**
	forced-choice	-613.46	-9.60	**< .001**	-0.699	[-0.893, -0.504]	**> 100**
	delayed forced-choice	-702.71	-9.08	**< .001**	-0.666	[-0.855, -0.478]	**> 100**
	written production	-782.02	-9.54	**< .001**	-0.668	[-0.853, -0.484]	**> 100**
	verbal-keyboard production	-877.35	-11.67	**< .001**	-0.696	[-0.867, -0.525]	**> 100**
	simple verbal production	-769.12	-12.26	**< .001**	-0.756	[-0.938, -0.573]	**> 100**
difficulty	verification	-1070.52	-13.60	**< .001**	-0.996	[-1.232, -0.761]	**> 100**
	forced-choice	-842.72	-15.46	**< .001**	-0.960	[-1.177, -0.743]	**> 100**
	delayed forced-choice	-1183.71	-15.95	**< .001**	-1.122	[-1.367, -0.877]	**> 100**
	written production	-1392.83	-17.41	**< .001**	-1.191	[-1.445, -0.936]	**> 100**
	verbal-keyboard production	-1278.02	-18.22	**< .001**	-1.014	[-1.224, -0.804]	**> 100**
	simple verbal production	-1223.22	-17.19	**< .001**	-1.202	[-1.458, -0.945]	**> 100**
distractor	forced-choice	-975.21	-8.83	**< .001**	-0.966	[-1.247, -0.685]	**> 100**
	delayed forced-choice	-46.17	-1.15	0.254	-0.042	[-0.115, -0.031]	0.00
	delayed forced-choice (second RT)	-15.11	-3.36	**.001**	-0.109	[-0.177, -0.041]	**66.87**

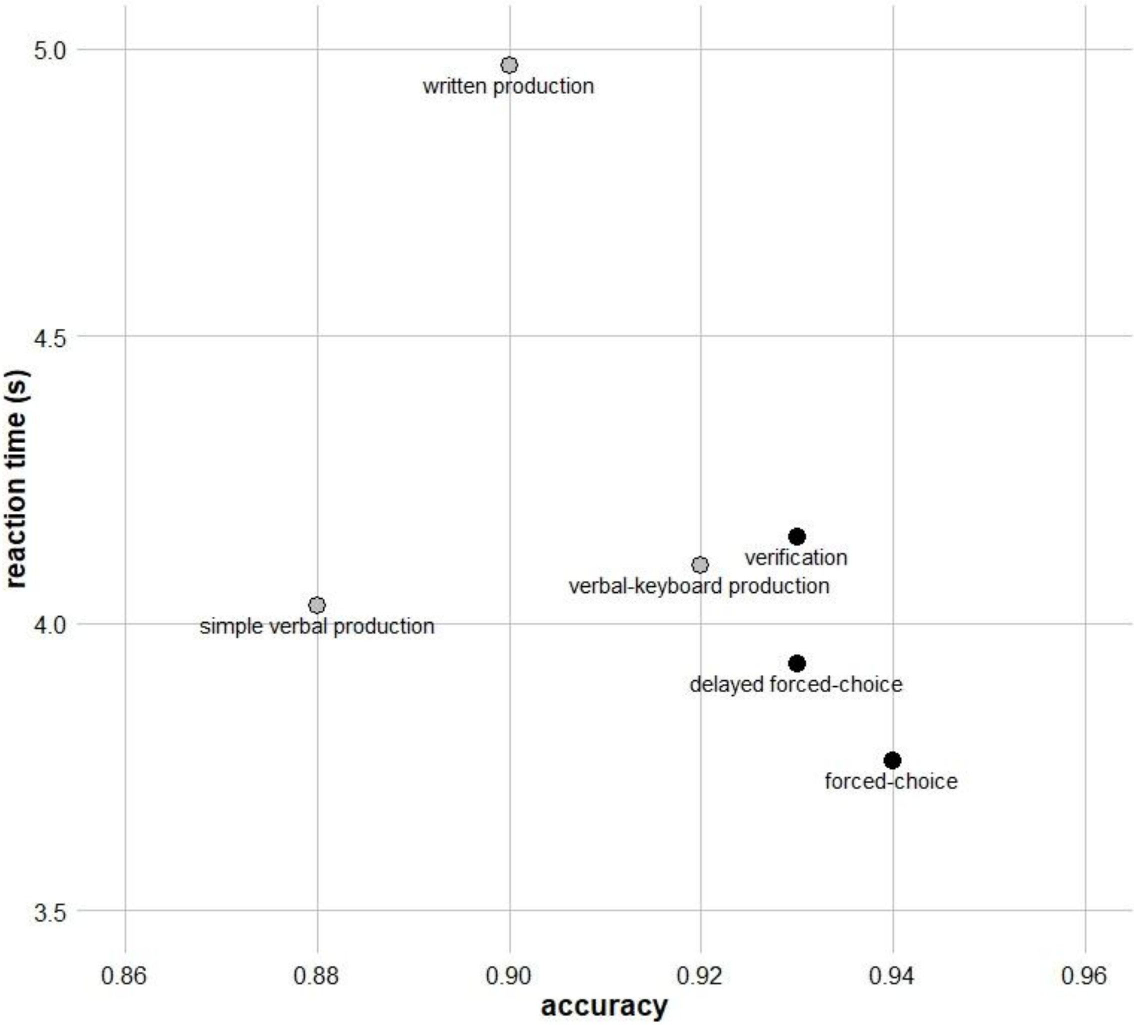
Notes. ΔM was reaction time in addition minus reaction time in subtraction (for operation), reaction time in simple minus reaction time in complex (for difficulty), and reaction time in distractor distance 2 minus reaction time in distractor distance 10 (for distractor). Bonferroni-Holm correction was applied to the pairwise comparisons.

Table S5 Pairwise comparisons of accuracy and reaction time between the carry effect and borrow effect

DV	paradigm	ΔM	*t*	*p*	Cohen's *d* [95% *CI*]		BF_{10}	BF_{01}
ACC	verification	-0.36	-1.37	.178	-0.178	[-0.434, 0.080]	0.34	2.92
	forced-choice	-0.31	-2.29	**.026**	-0.298	[-0.557, -0.036]	1.57	0.64
	delayed forced-choice	-0.02	-0.12	.902	-0.016	[-0.262, 0.231]	0.14	**7.19**
	written production	-0.21	-1.24	.221	-0.157	[-0.407, 0.094]	0.29	**3.49**
	verbal-keyboard production	0.07	0.39	.700	0.048	[-0.195, 0.291]	0.15	**6.84**
	simple verbal production	-0.31	-1.87	.066	-0.236	[-0.486, -0.015]	0.71	1.41
RT	verification	156	1.29	.203	0.168	[-0.090, 0.424]	0.31	**3.21**
	forced-choice	-16	-0.16	.877	-0.020	[-0.275, 0.235]	0.14	**6.94**
	delayed forced-choice	140	1.43	.158	0.180	[-0.069, 0.428]	0.36	2.76
	written production	84	0.67	.504	0.085	[-0.164, 0.334]	0.17	**5.80**
	verbal-keyboard production	233	2.24	**.029**	0.277	[0.029, 0.524]	1.38	0.72
	simple verbal production	60	0.55	.584	0.069	[-0.178, 0.316]	0.16	**6.27**

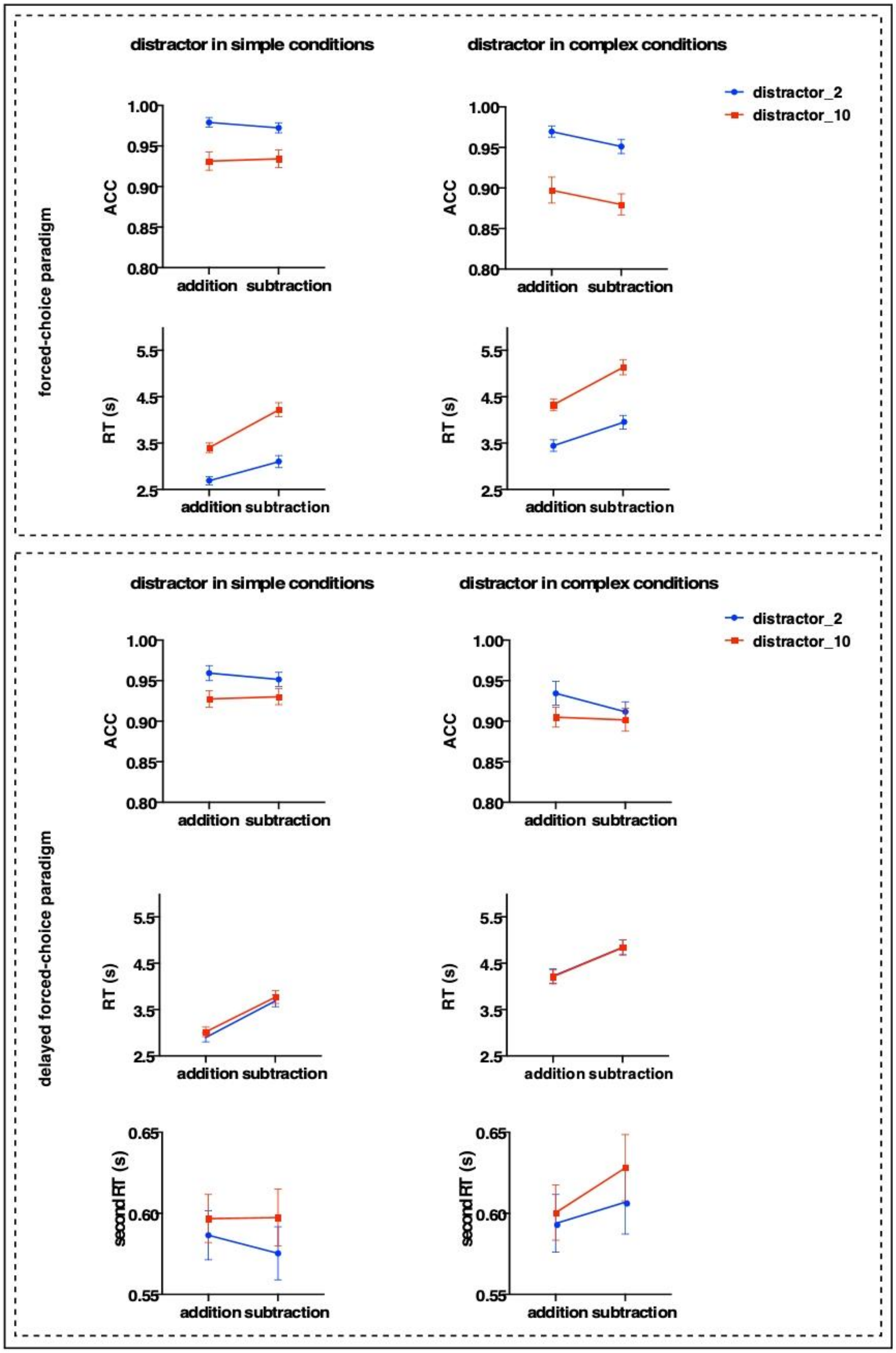
Notes. ΔM was logit-transformed accuracy of carry effect minus borrow effect and reaction time of carry effect minus borrow effect.

Figure S1. Summary of performance differences between paradigms



Notes. Results are based on the pairwise comparisons between all paradigms for logit-transformed accuracy and reaction time. For accuracy, the three decision paradigms (verification, forced-choice, delayed forced-choice) showed higher accuracy than the two production paradigms (written production, simple verbal production). Accuracy in verbal-keyboard production paradigm was similar to the delayed forced-choice paradigm while higher than in simple verbal production paradigm. For reaction time, written production showed longer reaction time than all other paradigms. Verification paradigm showed longer reaction time than the forced-choice paradigm. Verbal-keyboard production paradigm showed similar reaction times as the delayed forced-choice and simple verbal production paradigms.

Figure S2. Distractor effect for accuracy and reaction time (s) in (delayed) forced-choice paradigms



Notes. Accuracy and (second) reaction time were significantly longer when the distractor differed from the target by the decade position (±10) than by the unit position (±2) in (delayed) forced-choice paradigms, but not the first reaction time in the delayed forced-choice paradigm.

Table S6. Effect sizes for operation and difficulty effects across paradigms

		decision paradigms			production paradigms		
		verification	forced-choice	delayed forced choice	written production	verbal-keyboard production	simple verbal production
ACC	operation	0.095	0.231	0.159	0.177	0.374	0.319
	difficulty	0.246	0.430	0.365	0.319	0.538	0.467
RT	operation	-0.694	-0.699	-0.666	-0.668	-0.696	-0.756
	difficulty	-0.996	-0.960	-1.122	-1.191	-1.014	-1.202

Notes. For accuracy, the operation effect is only observed in verbal-keyboard production and simple verbal production paradigms, with larger effect size; the difficulty effect is observed in the five paradigms (forced-choice, delayed forced-choice, written production, verbal-keyboard production, and simple verbal production) except in the verification paradigm, with larger effect size in both verbal-keyboard production and simple verbal production paradigms. For reaction time, both operation and difficulty effects are observed in all six paradigms, with similar effect sizes in all six paradigms.

Table S7. Stimulus properties of the arithmetic task

	s-ds-1	s-ds-2	s-ds-3	s-ds-4	s-ds-5	s-ds-6	c-ds-1	c-ds-2	c-ds-3	c-ds-4	c-ds-5	c-ds-6
*a*_m	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83
*b*_m	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83	39.83
*r*_m	79.67	79.67	79.67	79.67	79.67	79.67	79.67	79.67	79.67	79.67	79.67	79.67
*a*_uni_m	4.42	4.00	3.58	3.17	2.75	3.17	6.50	6.50	6.50	6.50	6.50	6.50
*b*_uni_m	2.75	3.17	3.58	3.17	4.00	4.00	6.50	6.92	6.50	6.92	7.33	7.33
*r*_uni_m	7.17	7.17	7.17	6.33	6.75	7.17	3.00	3.42	3.00	3.42	3.83	3.83
*a*_dec_m	3.54	3.58	3.63	3.67	3.71	3.67	3.33	3.33	3.33	3.33	3.33	3.33
*b*_dec_m	3.71	3.67	3.63	3.67	3.58	3.58	3.33	3.29	3.33	3.29	3.25	3.25
*r*_dec_m	7.25	7.25	7.25	7.33	7.29	7.25	7.67	7.63	7.67	7.63	7.58	7.58
a>*b*/*a*< *b*	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12
uni-*a*>*b*/*a*< *b*	19/5	13/11	12/12	12/12	9/15	8/16	10/14	14/10	11/13	12/12	11/13	11/13
dec-*a*>*b*/*a*< *b*	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12	12/12
a-uni=1	2	4	5	5	9	6	0	0	0	0	0	0
b-uni=1	5	8	5	6	7	3	0	0	0	0	0	0
r-uni=1	0	0	0	0	0	0	6	4	6	7	3	3
a-uni=9	0	0	0	0	0	0	5	4	4	6	5	5
b-uni=9	0	0	0	0	0	0	2	5	3	6	8	9
r-uni=9	7	4	8	4	5	9	0	0	0	0	0	0

Notes. a, *b*, *r* represent the numbers in the math tasks in 6 datasets for two difficulty levels. In addition, *a* and *b* are the first and second addends and *r* is the sum. In subtraction, *r* is the minuend, *b* is the subtrahend, and *a* is the difference. "s-ds-1" means the simple condition in dataset 1, up to "c-ds-6" means the complex condition in dataset 6. "*a*_m", "*b*_m", and "*r*_m" means the mean of *a*, *b*, and *r* in each dataset. "*a*_uni_m", "*b*_uni_m", and "*r*_uni_m" represents the mean of the unit of *a*, *b*, and *r*. "*a*_dec_m", "*b*_dec_m", and "*r*_dec_m" represents the mean of the decade of *a*, *b*, and *r*. "*a*>*b*/*a*< *b*", "uni-*a*>*b*/*a*< *b*", and "dec-*a*>*b*/*a*< *b*" represents the portion of comparison between *a* and *b*, the unit of *a* and *b*, and the decade of *a* and *b*. "*a*-uni=1" indicates that the unit of *a* is equal to 1, "*a*-uni=9" indicates that the unit of *a* is equal to 9, so as *b* and *r*.

Table S8a. Split-half reliability (corrected for attenuation): correlations within each paradigm

	accuracy				reaction time			
	addition simple	addition complex	subtraction simple	subtraction complex	addition simple	addition complex	subtraction simple	subtraction complex
verification	.230	.079	.117	.154		.716***	.726***	.753***
forced-choice	.056	.682***	.164	.112	.717***	.682***	.791***	.766***
delayed forced-choice	.259*	.239	.272*	.414***	.824***	.239	.790***	.855***
written production	.136	.399***	-.010	.333***	.759***	.820***	.797***	.856***
verbal-keyboard production	.612***	.753***	.630***	.713***	.966***	.933***	.965***	.973***
simple verbal production	.602***	.818***	.250*	.513***	.896***	.817***	.751***	.759***

Notes. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S8b. Construct validity: correlations across paradigms for mean reaction time

	verification	forced-choice	delayed forced-choice	written production	verbal-keyboard production	simple verbal production
verification						
forced-choice	.794***					
delayed forced-choice	.740***	.788***				
written production	.618***	.815***	.815***			
verbal-keyboard production	.705***	.796***	.798***	.818***		
simple verbal production	.678***	.720***	.774***	.722***	.773***	

Notes. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S8c. Construct validity: correlations across paradigms for mean accuracy

	verification	forced-choice	delayed forced-choice	written production	verbal-keyboard production	simple verbal production
verification						
forced-choice	.439***					
delayed forced-choice	.548***	.507***				
written production	.565***	.592***	.728***			

verbal-keyboard production	.363**	.435***	.633***	.555***	
simple verbal production	.324*	.297*	.306*	.499***	.324*

Notes. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S9a. Reaction time comparisons between verification and forced-choice paradigms

						95% *CI* for Cohen's *d*	
		t	*df*	*p*	Cohen's *d*	Lower	Upper
V_as	FC_as_2	4.74	64	**< .001**	0.587	0.322	0.849
	FC_as_10	-2.50	64	**.015**	-0.310	-0.558	-0.060
V_ac	FC_ac_2	6.48	64	**< .001**	0.804	0.521	1.081
	FC_ac_10	0.12	64	.906	0.015	-0.228	0.258
V_ss	FC_ss_2	7.19	64	**< .001**	0.891	0.601	1.177
	FC_ss_10	-1.36	64	.180	-0.168	-0.412	0.077
V_sc	FC_sc_2	6.81	64	**< .001**	0.844	0.558	1.126
	FC_sc_10	-0.55	64	.585	-0.068	-0.311	0.176

Notes. "V" is the abbreviation of verification, "FC" is the abbreviation of forced-choice. "as" represents addition simple, which doesn't involve carry operation; "ac" represents addition complex, which involves carry operation; "ss" represents subtraction simple, which doesn't involve borrow operation; "sc" represents subtraction complex, which involves borrow operation. "_2" represents a distractor with "± 2", "_10" represents a distractor with "± 10". $p < .05$ are marked in bold.

Table S9b. Accuracy comparisons between verification and forced-choice paradigms

						95% *CI* for Cohen's *d*	
		t	*df*	*p*	Cohen's *d*	Lower	Upper
V_as	FC_as_2	-3.63	64	**< .001**	-0.451	-0.704	-0.194
	FC_as_10	0.22	64	.824	0.028	-0.216	0.271
V_ac	FC_ac_2	-3.87	64	**< .001**	-0.479	-0.735	-0.221
	FC_ac_10	2.23	64	**.029**	0.277	0.028	0.523
V_ss	FC_ss_2	-2.54	64	**.014**	-0.315	-0.563	-0.064
	FC_ss_10	1.03	64	.306	0.128	-0.117	0.372
V_sc	FC_sc_2	-2.52	64	**.014**	-0.313	-0.561	-0.063
	FC_sc_10	1.26	64	.214	0.156	-0.089	0.400

Notes. "V" is the abbreviation of verification, "FC" is the abbreviation of forced-choice. "as" represents addition simple, which doesn't involve carry operation; "ac" represents addition complex, which involves carry operation; "ss" represents subtraction simple, which doesn't involve borrow operation; "sc" represents subtraction complex, which involves borrow operation. "_2" represents a distractor with "± 2", "_10" represents a distractor with "± 10". $p < .05$ are marked in bold.

A2. Study 2

A2.1 Publication of Study 2

Yao, X., Huber, J. F., Li, Z., Findik, Y., Nuerk, H.-C., & Artemenko, C. (2026). The dynamics of state math anxiety vary by paradigm and timing during arithmetic. *npj* Science of Learning. Advance online publication. https://doi.org/10.1038/s41539-025-00398-z

npj | science of learning

Article

Published in partnership with The University of Queensland

https://doi.org/10.1038/s41539-025-00398-z

The dynamics of state math anxiety vary by paradigm and timing during arithmetic

Check for updates

Xinru Yao[1,2], Julia F. Huber[1], Zixu Li[3], Yaren Findik[1], Hans-Christoph Nuerk[1,2,4] & Christina Artemenko[1,2,4]

Math anxiety impairs performance, but how its state and trait components interact with task characteristics remains unclear. We examined how state math anxiety varies as a function of trait math anxiety, task paradigm, and temporal dynamics, and how trait math anxiety relates to arithmetic performance. Results revealed that production paradigms, which require generating answers, elicited higher state math anxiety compared to decision paradigms, particularly for individuals with high trait math anxiety. Looking into different task phases, state math anxiety decreased during arithmetic due to habituation and after arithmetic due to relief. Additionally, the anxiety-complexity effect was replicated: Individuals with higher trait math anxiety were slower in solving complex arithmetic with a carry or borrow operation. This study confirmed the situation-dependent characteristics of state math anxiety and its dependency on paradigm and trait math anxiety, with implications for designing interventions that mitigate anxiety and optimize learning.

Mathematics is a fundamental subject in education and an essential skill for daily life. From science and engineering to finance and technology, mathematical skills are needed for numerous fields, making proficiency in math crucial for academic and professional success. Studies have consistently shown that math skills are linked to problem-solving skills, logical thinking[1,2], and career opportunities, especially in STEM fields[3]. However, despite its importance, mathematics often provokes negative emotions, particularly math anxiety, which can significantly influence performance and long-term academic outcomes due to avoidance of the subject.

Math anxiety, characterized by tension and fear that interfere with manipulating numbers and solving mathematical problems[4], has long been a topic of significant interest in both psychological and educational research. Math anxiety has consistently been linked to poor math performance[5,6]. Moreover, an anxiety-complexity effect was found, indicating that math-anxious individuals have difficulties particularly with more complex arithmetic[4,7,8]. This means that task difficulty affects performance in individuals with higher levels of anxiety more than in individuals with lower levels of anxiety. Several theoretical accounts have been proposed to explain the relationship between math anxiety and performance, which can be broadly categorized into cognitive disruption explanations, competency-based explanations, and integrative approaches.

The Disruption Account[9,10] postulates that math anxiety triggers intrusive thoughts and ruminations, which occupy working memory resources, reduce cognitive efficiency and subsequently impair math performance. This explains the anxiety-complexity effect, as more complex arithmetic (such as addition with carry operation and subtraction with borrow operation) relies on working memory[11] and thus is even more impaired than performance in simple arithmetic. However, recent findings from network analysis suggest that math anxiety and working memory are independently linked to math performance, indicating that the relationship between math anxiety and performance is not solely dependent on working memory[12]. Further support for the Disruption Account is provided by the Processing Efficiency Theory and Attentional Control Theory. The Processing Efficiency Theory[13] suggests that anxiety reduces the efficiency (not effectiveness) of cognitive processing by transferring attentional resources to task-irrelevant worry, thereby leaving fewer resources available for task performance. The Attentional Control Theory[14] further emphasizes that anxiety disrupts the balance between goal-directed (top-down) and stimulus-driven (bottom-up) attentional control, leading to impaired concentration and thus impairing task performance.

While the Disruption Account emphasizes how anxiety impairs performance through cognitive mechanisms, the Reduced Competency Account[10] offers an alternative perspective, suggesting that math anxiety is the result of poor math ability, where reduced competency leads to difficulties in learning and performance, ultimately causing anxiety, with individuals often avoiding math-related tasks and opportunities for improvement[15–17]. An extreme example of this is that children with dyscalculia exhibit higher levels of math anxiety compared to typically

[1]Department of Psychology, University of Tuebingen, Tuebingen, Germany. [2]LEAD Graduate School and Research Network, University of Tuebingen, Tuebingen, Germany. [3]Department of Social Work and Social Administration, University of Hong Kong, Hong Kong, China. [4]German Centre for Mental Health (DZPG), Tuebingen, Germany. e-mail: xinru.yao.tuebingen@gmail.com; hc.nuerk@uni-tuebingen.de

developing children[18]. Extending the Reduced Competency Account, the Interpretation account[10] argues that math anxiety stems not only from poor math skills or negative experiences but from how individuals interpret and appraise their math-related experiences.

Rather than viewing these accounts as mutually exclusive, all different theoretical accounts can be integrated in a reciprocal theory, stating a bidirectional relationship between math anxiety and math performance. Accordingly, math anxiety and mathematics performance influence each other resulting in a vicious cycle[19,20]. Taken together, math anxiety shapes and is shaped by performance and experiences in math.

Based on the state-trait anxiety model[21], two distinct forms of math anxiety can be differentiated[22]: On the one hand, trait math anxiety refers to a stable, enduring personality trait reflecting a general tendency to feel anxious during math-related tasks. On the other hand, state math anxiety refers to a temporary, situation-specific form of anxiety that arises in response to math tasks. State math anxiety can be accompanied by physiological responses such as heightened autonomic nervous system arousal[20,22]. Crucially, state math anxiety represents the dynamic, moment-to-moment emotional experience during actual mathematical engagement, making it particularly relevant for understanding real-time cognitive processing and performance.

The relationship between trait and state math anxiety is complex. Individuals with high trait math anxiety are more likely to experience high state anxiety during math tasks[23], though the two forms of anxiety are shaped by different factors. Trait math anxiety is influenced by past experiences and long-term beliefs about math, while state math anxiety is sensitive to immediate contexts and specific task characteristics[24–27]. Task difficulty can be considered as a situational determinant of state math anxiety, with more difficult math problems being associated with higher levels of state anxiety in children[27,28] and adults[24,29,30]. The interplay between state and trait math anxiety and their combined impact on real-time math performance is a central question. Some studies propose that state math anxiety directly mediates the negative effect of trait math anxiety on performance, particularly under demanding conditions[31]. However, Pelegrina et al.[32] found that when both types of math anxiety were examined simultaneously, trait math anxiety was a stronger predictor of math performance than state math anxiety, with the effects of state math anxiety largely attributable to shared variance with trait math anxiety.

Therefore, to accurately assess math anxiety during arithmetic tasks, it is essential to consider both trait and state math anxiety. Trait assessments of math anxiety are designed to measure traits, i.e., stable predispositions of an individual shown across many situations, but not situational fluctuations. As trait assessments are often based on hypothetical or retrospective questionnaires, there might be a tendency to overestimate the anxiety experienced in mathematical situations due to the intensity bias[33–35] and a bias due to beliefs about one's own competence in math[34]. In contrast, state questionnaires rather directly assess emotions in a certain situation and thus provide a more accurate reflection by capturing emotions in real-time. Furthermore, state math anxiety can be captured through different modalities, such as self-report and physiological measures, which may provide distinct information and relate differently to trait measures[30,31]. This leads to discrepancies between trait and state math anxiety[34,36] and highlights the value of state assessments for math anxiety[24].

In distinguishing between trait and state math anxiety, other individual differences – such as test anxiety, math self-concept, math ability, working memory capacity, and gender – may also play a role[24,37]. However, the present study focuses primarily on state math anxiety and its dependance on trait math anxiety.

As state math anxiety is sensitive to situational fluctuations, the way math performance is assessed might impact state math anxiety. Math performance can be assessed in several ways: Children in school almost exclusively solve mental arithmetic in production paradigms with an open answer format, whereas studies in laboratory settings often use decision paradigms with given answers. Studies have shown that children's performance can vary depending on the response format, with better performance observed in decision paradigms compared to production paradigms[38,39]. Similarly, an experiment in adults showed that decision paradigms (e.g., verification, forced-choice, and delayed forced-choice) lead to better performance compared to production paradigms (e.g., written production, verbal-keyboard production, and simple verbal production)[40]. Thus, the response format of an arithmetic task creates different situations so that performance varies with paradigm.

Moreover, math performance is related to math anxiety. This raises the question of whether (state) math anxiety also varies across different paradigms. Given that state math anxiety is sensitive to the specific situational context, it is plausible that varying paradigms, each involving distinct solution processes and different levels of uncertainty about mistakes and failures, may impact state math anxiety and its relationship to performance. The open format of response and the generally lower performance observed in production paradigms may lead to more anxiety compared to decision paradigms where possible responses are given, as production paradigms place higher calculation demands and increase uncertainty about correctness. Therefore, the first objective of this study is to evaluate whether state math anxiety depends on the paradigm.

Additionally, the relationship between speed and accuracy can vary with math anxiety levels: compared to individuals with low math anxiety, who are fast and accurate in arithmetic, individuals with moderate math anxiety are slower and individuals with high math anxiety are less accurate[4]. Consequently, there is a need to investigate whether specific speed-accuracy trade-offs can be observed due to math anxiety. The speed-accuracy trade-off is a strategic adjustment in the decision process that adapts to environmental demands[41,42]. As math-anxious individuals may prioritize speed or accuracy differently depending on the paradigm, potentially influencing their overall performance, the potential for speed-accuracy trade-offs needs to be explored.

When investigating state math anxiety, it is important to account for possible fluctuations throughout a mathematical task. Research shows that state anxiety anticipated before a math task can negatively affect performance. For example, Orbach and Fritz[43] found that children's math performance was impacted by state anxiety before – but not after – completing math tasks[44,45]. Similarly, Goetz et al.[46] observed higher levels of state anxiety before a test compared to afterwards, suggesting that the anticipation of the task plays a significant role. Taken together, these results suggest that state math anxiety is higher when anticipating a math task compared to the relief after completing the math task.

Conversely, Conlon et al.[47] found that state math anxiety could increase after the math task, particularly after challenging problem-solving tasks, reflecting the cognitive demands and complexity of the task. Supporting this evidence, physiological markers such as heart rate and skin conductance revealed that state math anxiety increased during a math exam, especially in later stages, likely due to rising time pressure and task difficulty[48]. These seemingly contradictory findings highlight the complexity of state math anxiety and its temporal dynamics, raising the question: How does the level of state math anxiety change during math tasks? Is it elevated or reduced compared to the pre-task level?

Resolving these inconsistencies requires a clearer understanding of how specific phases of a task (i.e., pre-task anticipation, mid-task progression, and post-task evaluation) contribute to state math anxiety. Research on anxiety therapies, such as exposure techniques, shows that state anxiety usually decreases over time due to the process of habituation[49]. However, this general decrease over time overlaps with the mid-task and post-task phases, making it unclear what exactly causes the reduction in anxiety: Is it only the relief after task completion that reduces anxiety, or rather the repetitive exposure to the task? This is another objective of the current study. To distinguish whether these temporal patterns are specific to math anxiety or reflect broader anxiety dynamics during task performance, the present study also assessed general state anxiety.

The present study aims to provide a comprehensive understanding of the dynamics of math anxiety and its impact on arithmetic performance, considering both paradigm effects and temporal trends. Specifically, we

Table 1 | Descriptive statistics of state (math) anxiety measures across task phases

Category	Temporal progression	State math anxiety	State anxiety
three task phases	pre-task	2.06 (1.05)	1.82 (0.75)
	mid-task	1.66 (0.69)	1.61 (0.57)
	post-task	1.46 (0.69)	1.29 (0.51)
mid-task measurement time	1	2.01 (0.76)	1.84 (0.79)
	2	1.84 (0.86)	1.71 (0.74)
	3	1.86 (0.92)	1.76 (0.82)
	4	1.63 (0.79)	1.54 (0.74)
	5	1.49 (0.68)	1.42 (0.62)
	6	1.50 (0.69)	1.42 (0.61)
paradigms	decision	1.56 (0.61)	1.45 (0.54)
	production	1.88 (0.76)	1.78 (0.64)

Notes. Values are presented as *M* (*SD*) for each measure. State math anxiety was measured with the 1-item version (SMA-1) before and after the arithmetic task.

examined how state math anxiety varies as a function of trait math anxiety, depending on task paradigm and on time, respectively. Additionally, we further evaluated the relation of trait math anxiety to performance in this context. The following conceptual hypotheses were preregistered (https://aspredicted.org/gf8dc.pdf) before data collection:

For math anxiety differences between paradigms (H1), (H1a) state math anxiety across paradigms (confirmatory)—we expect state math anxiety to be higher in production paradigms, where participants must generate the answer themselves, compared to decision paradigms, where the correct answer is selected from given options. Additionally, higher trait math anxiety is expected to be associated with a larger difference in state math anxiety between the paradigms (interaction between paradigm and trait math anxiety). (H1b) trait math anxiety and arithmetic performance (confirmatory & exploratory)—we expect a negative relation between trait math anxiety and arithmetic performance, with higher math anxiety associated with longer response times and lower accuracy. Additionally, we will explore whether the relation between trait math anxiety and arithmetic performance differs between production and decision paradigms.

For temporal dynamics of state math anxiety (H2), (H2a) changes across phases of the task (exploratory)—we will explore changes in state math anxiety across different phases of an arithmetic task (pre-, mid-, post-task), i.e., how state math anxiety changes from before to during and after arithmetic. Additionally, we will examine how trait math anxiety influences these changes in state math anxiety across the task phases. (H2b) detailed dynamics during the task (exploratory)—we will further explore the temporal dynamics of state math anxiety in more detail during the arithmetic task, using six measurement time points taken during breaks at the midpoint of each paradigm block. This detailed analysis is based on an additional related preregistration (https://aspredicted.org/sc87-3ntj.pdf), which was completed after data collection but before any data inspection or analysis regarding this hypothesis.

Regarding the relation between trait math anxiety and performance (H3), (H3a) anxiety-complexity effect (confirmatory)—an anxiety-complexity effect is expected regardless of the paradigm, i.e., higher levels of trait math anxiety are associated with larger carry/borrow effects. (H3b) speed-accuracy trade-off (exploratory)—for trait math anxiety, we will further explore potential speed-accuracy trade-offs within and between subjects in the different paradigms.

The operationalization of the study considers measures of trait and state math anxiety while participants are performing an arithmetic task. Trait math anxiety (AMAS) was assessed only before the arithmetic task, while state math anxiety (SMA) and state anxiety (STAI-SKD) were assessed before, during and after the arithmetic task. The arithmetic task consisted of two-digit addition and subtraction problems presented in different paradigms (decision vs. production). Arithmetic performance outcomes were accuracy (ACC) and response time (RT).

Results

For data analysis, paradigm effects (H1) and temporal dynamics (H2) were examined in separate models as they address conceptually distinct research questions. Results for the anxiety-complexity effect (H3a) and speed-accuracy trade-off (H3b) can be found in Table S2. Participants' mean trait math anxiety was 1.98 ($SD = 0.68$). Descriptive analysis results were shown in Table 1 and Table S1.

Paradigm-dependent analysis for state math anxiety

Regarding H1a, an LMM with state math anxiety as the dependent variable was conducted including fixed effects for paradigm (production vs. decision), trait math anxiety, and their interaction. The LMM further included a random intercept for subject (but not – as incorrectly preregistered – for item, because state measures were not assessed at an item level but only at a block level). Additionally, the LMM did not include a random slope for paradigm, as specified in the preregistration, because incorporating this would have made the random effects structure too complex given the limited number of observations.

The final LMM (Model A) included fixed effects for paradigm, trait math anxiety and their interaction, with a random intercept for subject (see Table 2, Fig. 1a). The main effect of paradigm indicates that state math anxiety is higher in production paradigms compared to decision paradigms, with an estimated increase of 0.29. The main effect of trait math anxiety indicates that state math anxiety increases with increasing trait math anxiety, with an estimate of 0.28. The interaction of paradigm and trait math anxiety indicates that the effect of trait math anxiety on state math anxiety is larger in production than in decision paradigms, by an estimate of 0.24. These results imply that state math anxiety is higher in production than in decision paradigms, particularly for individuals with higher trait math anxiety.

Paradigm-dependent analysis for arithmetic performance

Regarding H1b, an LMM with RT as the dependent variable and a GLMM with ACC as the dependent variable were conducted including fixed effects for paradigm, trait math anxiety, and their interaction. The (G)LMM further included random intercepts for both subject and item as well as a random slope for paradigm.

The final LMM for RT (Model B) included fixed effects for trait math anxiety and paradigm, with random intercepts for subject and item as well as a random slope for paradigm (see Table 2, Fig. 1b). The main effect of paradigm indicates that arithmetic in production paradigms needs longer by an estimate of 0.41 s to be solved than in decision paradigms. The main effect of trait math anxiety indicates that for every unit increase in trait math anxiety, the response time increases by an estimate of 0.36 s, so that individuals with higher trait math anxiety take longer to solve arithmetic.

Similar to RT, the final GLMM for ACC (Model C) included fixed effects for trait math anxiety and paradigm, with random intercepts for subject and item as well as a random slope for paradigm (see Table 2, Fig. 1c). The main effect of paradigm indicates that accuracy was lower by an estimate of -0.28 in production compared to decision paradigms. The main effect of trait math anxiety was marginally significant with an estimate of -0.21, indicating that individuals with higher math anxiety tend to make more errors in arithmetic.

Together, the results suggest that individuals with higher trait math anxiety show worse arithmetic performance, and production paradigms are more difficult than decision paradigms.

Time analysis pre-, mid-, and post-arithmetic task

Regarding H2a, LMMs with state math anxiety (SMA-1) and state anxiety as dependent variables were conducted including fixed effects

Table 2 | LMM/GLMM results

Predictors	β	*CI*	*t / z*	*p*	R^2
Model A: LMM for paradigm and trait math anxiety on state math anxiety					0.87
(intercept)	1.57	1.42 – 1.72	20.74	**< 0.001**	
paradigm	0.29	0.21 – 0.38	6.64	**< 0.001**	
trait math anxiety	0.28	0.06 – 0.50	2.52	**0.013**	
paradigm × trait math anxiety	0.24	0.11 – 0.37	3.64	**< 0.001**	
Model B: LMM for trait math anxiety and paradigm on response time					0.42
(intercept)	3.96	3.75 – 4.16	37.58	**< 0.001**	
paradigm	0.41	0.30 – 0.52	7.43	**< 0.001**	
trait math anxiety	0.36	0.06 – 0.65	2.38	**0.017**	
paradigm × trait math anxiety					
Model C: GLMM for trait math anxiety and paradigm on accuracy					0.15
(intercept)	2.68	2.53 – 2.82	35.42	**< 0.001**	
paradigm	-0.28	-0.40 – -0.16	-4.46	**< 0.001**	
trait math anxiety	-0.21	-0.41 – 0.00	-1.93	0.053	
paradigm × trait math anxiety					
Model D: LMM for three-time points analysis on state math anxiety					0.70
(intercept)	1.73	1.59 – 1.87	23.95	**< 0.001**	
time	-0.30	-0.39 – -0.22	-7.32	**< 0.001**	
trait math anxiety	0.60	0.39 – 0.81	5.63	**< 0.001**	
time × trait math anxiety	-0.31	-0.43 – -0.19	-5.07	**< 0.001**	
time^2					
time^2 × trait math anxiety					
Model E: LMM results for six-time points analysis on state math anxiety					0.75
(intercept)	1.72	1.48 – 1.96	14.03	**< 0.001**	
time	-0.11	-0.13 – -0.08	-8.96	**< 0.001**	
trait math anxiety	0.40	0.18 – 0.62	3.65	**< 0.001**	
time × trait math anxiety	-0.07	-0.11 – -0.04	-3.96	**< 0.001**	
time^2					
time^2 × trait math anxiety					

Notes. All models shown here are reduced models. *t* value for LMM and *z* value for GLMM. Bold values indicate statistical significance. Time and trait math anxiety were centered. Paradigm was dummy coded with decision paradigm as reference for production paradigm. Conditional R^2 quantifies the proportion of variance explained by the entire model, including both fixed and random effects. Excluded factors from the full models are shown in italics. Full model specifications and model selection procedures are detailed in Table S3.

for both the linear (time) and quadratic term of time (time^2), trait math anxiety, the interaction between time and trait math anxiety, and the interaction between time^2 and trait math anxiety. The linear term of time includes 3 time points: pre-, mid- (average across 6 measurement points) and post-arithmetic task. The quadratic term of time was introduced to account for non-linear trajectories, for example, an initial increase in anxiety during the task followed by a decrease at the end after the task is successfully completed. The LMMs further included a random intercept for subject.

The final LMM model on state math anxiety (Model D) included fixed effects for time, trait math anxiety, and the interaction between time and trait math anxiety, with a random intercept for subject (see Table 2, Fig. 2a & b). The main effect of time indicates that state math anxiety decreases over time by an estimate of -0.30 per time point. The main effect of trait math anxiety indicates that for each unit increase in trait math anxiety, state math anxiety increases by an estimate of 0.60, so that individuals with higher trait math anxiety show higher levels of state math anxiety. The interaction effect of time and trait math anxiety with an estimate of -0.31 indicates that the decrease in state math anxiety over time is stronger for individuals with higher trait math anxiety. A similar analysis was conducted with state anxiety as a dependent variable (for results see Fig. S1).

Time analysis during the arithmetic task

Regarding H2b, LMMs with state math anxiety (SMA) and state anxiety as dependent variables were conducted including fixed effects for time (6 measurement times during arithmetic), the quadratic term of time (time^2), trait math anxiety, the interaction between time and trait math anxiety, and the interaction between time^2 and trait math anxiety. The LMMs further included random intercepts for subject and paradigm. Paradigm was included as a random effect to account for paradigm-specific effects.

The final LMM model (Model E) included fixed effects for time, trait math anxiety, the interaction between the time and trait math anxiety, with random intercepts for subject and paradigm (see Table 2, Fig. 2c & d). The main effect of time indicates that for each repetition of the arithmetic task, state math anxiety decreases over time by an estimate of -0.11. The main effect of trait math anxiety indicates that for each unit increase in trait math anxiety, state math anxiety increases by an estimate of 0.40, so that individuals with higher trait math anxiety show higher levels of state math anxiety. The interaction of time and trait math anxiety with an estimate of -0.07 indicates that, as time progresses, individuals with higher trait math anxiety tend to experience a decrease in state math anxiety at a slightly faster rate compared to those with lower trait math anxiety. This suggests that while initial

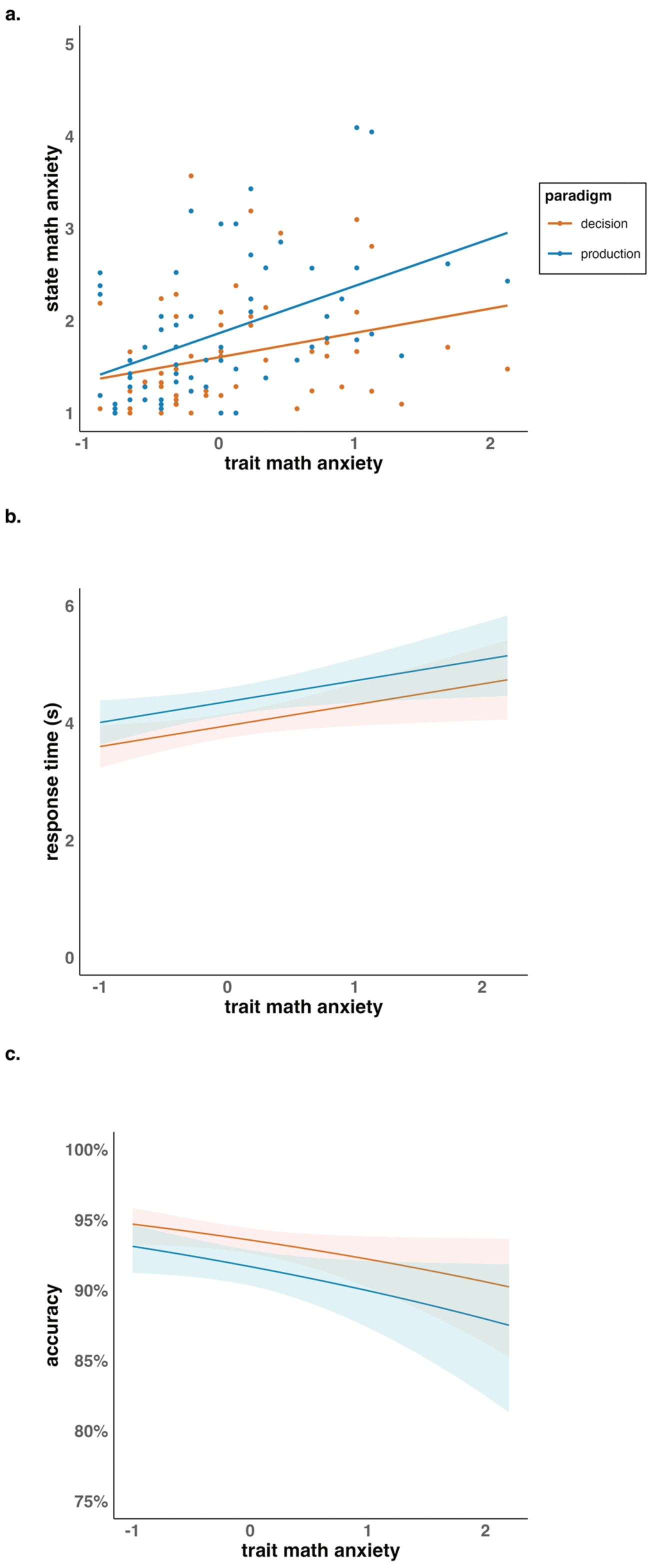


Fig. 1 | Relation of trait math anxiety to state math anxiety and performance dependent on paradigm. a shows the interaction between paradigm and trait math anxiety on state math anxiety. **b**, **c** show the effects of paradigm and trait math anxiety on performance in terms of response time and accuracy, respectively. Response time analyses were based on correctly solved trials only.

anxiety levels may be elevated, individuals with higher trait math anxiety may adapt or regulate their state math anxiety over the progress of the task. A similar analysis was conducted with state anxiety as dependent variable (for results see Fig. S1).

Discussion

This study explored the dynamics of state math anxiety across different paradigms and its relation to trait math anxiety and arithmetic performance. The findings highlight three key insights: (1) Production paradigms elicit higher state math anxiety than decision paradigms, especially for individuals with higher levels of trait math anxiety. (2) Higher trait math anxiety is associated with worse performance, especially in production paradigms. (3) State math anxiety decreases over time from before to after the arithmetic task as well as during the arithmetic task, especially for individuals with higher trait math anxiety who start with higher levels of state math anxiety, suggesting habituation and regulation mechanisms in managing anxiety.

The response format in math tasks indeed influences emotional states. In line with the interaction model of anxiety[50], situational characteristics such as paradigm can modulate momentary (state) anxiety. Higher state math anxiety in production paradigms compared to decision paradigms suggests that generating answers, rather than choosing from answers, imposes greater emotional demands. In decision paradigms, performance is better and various strategies can be used to solve arithmetic[40]. For instance, estimation strategies, verifying the correctness of a given answer, or rejecting incorrect answers contribute to more efficient and accurate task performance[51,52]. Better task performance was also shown to be associated with lower state math anxiety[24], so that the paradigm effect on state math anxiety might be driven by performance differences. In contrast, production paradigms require individuals to generate answers without the aid of pre-existing choices, like in real world contexts. This demands a higher cognitive load and is associated with less security, leading to higher state math anxiety, as found in the current study. In sum, production paradigms in arithmetic increase difficulty and state math anxiety compared to decision paradigms.

Notably, the production-induced increase in state math anxiety was predicted by trait math anxiety. Individuals with higher trait math anxiety are particularly sensitive to the cognitive demands imposed by the production paradigm, and thus experience especially more state math anxiety when they are required to exactly calculate the result without checking preexisting solutions. Previous research[7,53] has primarily demonstrated that high trait math anxiety reduces performance, particularly under complex task conditions, reflecting the anxiety-complexity effect, which is replicated in the current study[4,7,8]. Our findings show that state math anxiety is similarly affected by the interplay of trait math anxiety and complex task demands (by paradigm). However, state math anxiety might be rather an accompanying phenomenon for affective experiences in a situation (as measured by state math anxiety) than a factor that further impairs performance[24].

Taken together, state math anxiety was found to vary with the paradigm in which arithmetic needed to be solved, likely due to the immediate situational demands of the task, such as the higher difficulty of producing compared to selecting an answer. Trait math anxiety moderated the paradigm effect on state math anxiety, with individuals higher in trait math anxiety showing larger differences in state math anxiety between paradigms. This pattern is consistent with theoretical models suggesting that trait anxiety influences the magnitude of state anxiety responses to situational demands[21,22,24], pointing at the directive role of trait anxiety in generating state anxiety in specific situations. These findings have significant implications for educational settings, suggesting that the response format might be relevant in either mitigating or amplifying anxiety during mathematics for students with high levels of trait anxiety.

Complementary exploratory analyses further suggest that the heightened state math anxiety observed in production paradigms may primarily stem from fear of making mistakes rather than from increased response times. Specifically, higher levels of state math anxiety were associated with lower accuracy in production compared to decision paradigms but not with response time (Table S4). This pattern indicates that emotional tension may arise from concerns about correctness and error likelihood in open response formats rather than from time pressure.

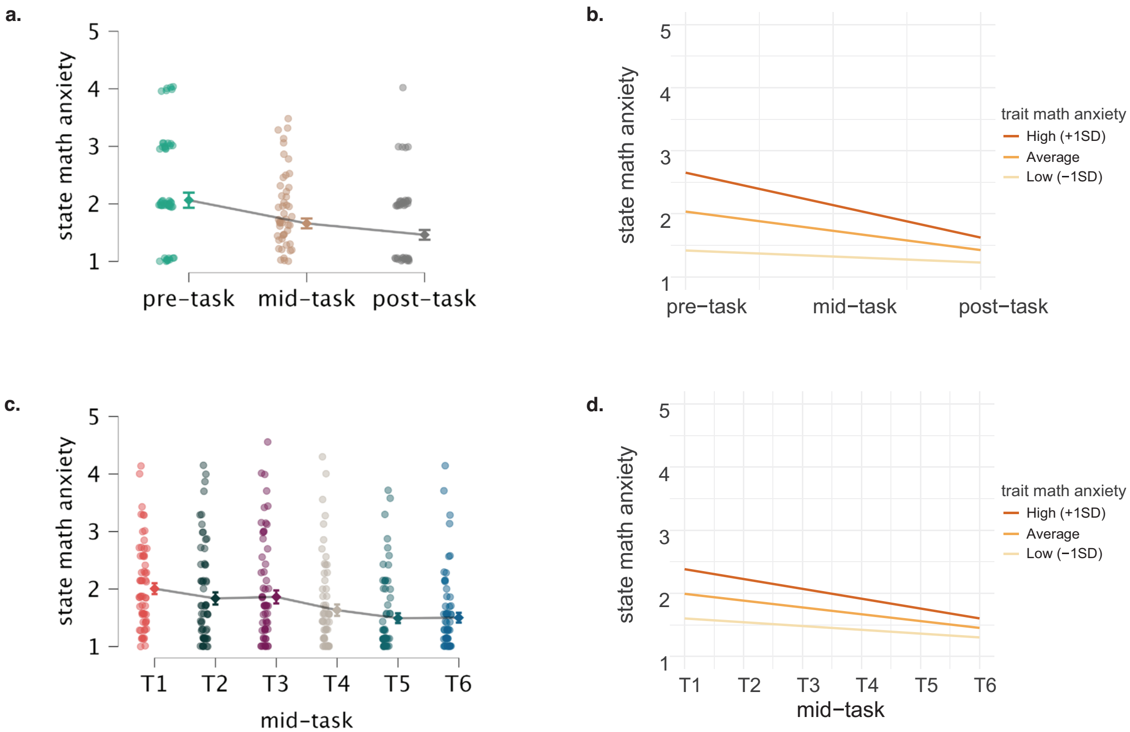


Fig. 2 | State math anxiety changes across different task phases and measurement times. **a**, **c** show the decrease in state math anxiety across the three task phases and across the six measurement times during the arithmetic task, respectively. Error bars represent the standard error of the mean (*SEM*). **b**, **d** show a simple slope analysis for state math anxiety depending on trait math anxiety (average level, high level with 1 *SD* above the average, and low level with 1 *SD* below the average) across the three task phases and across the six measurement times during the arithmetic task. State math anxiety was measured via questionnaires and therefore reflects all participants regardless of their individual performance.

The temporal analysis of state math anxiety adds another layer of depth to our understanding of how anxiety changes over time across different phases of mathematical tasks. State math anxiety was found to be higher before the arithmetic task, replicating results from Goetz et al.[46], decreased during the task and finally was lower after the task, especially for individuals with high trait math anxiety. This suggests that while these individuals initially experience increased expectation anxiety, their engagement with the task may activate emotion regulation mechanisms that contribute to the reduction in state math anxiety as they progress and afterwards, they are relieved that the task is over. Alternatively, the drop in anxiety during and after the task may also reflect participants developing a more realistic perception of the actual task demands, with anticipatory anxiety inflated by generalized math fears being recalibrated once they confront the true difficulty. The findings align with research showing that activation of frontal brain regions during the anticipation phase can support cognitive control and emotion regulation, mitigating performance deficits[45].

The multidimensional interaction model of anxiety[54] posits that state anxiety is influenced by interactions between trait anxiety and situational factors. Individuals with higher trait anxiety may have increased state anxiety at the start of a math task compared to individuals with lower trait math anxiety. Therefore, they need to regulate their emotions and thus anxiety decreases as they become familiar with the task. Consequently, this reduction is not solely due to habituation but depends on the individual predispositions and the characteristics of the task. Our results revealed similar linear decreases in state math anxiety for individuals with higher trait math anxiety from pre to post as well as during the task. Moreover, the time-wise correlations (Table S5 and Figure S3) revealed that the association between state math anxiety and performance was strongest at the beginning of the experiment and gradually weakened over time (for response time). Thus, both (successfully) completing the task (pre-post task comparison) and habituation over time (repetitions of the arithmetic task in different paradigms) play a role in the dynamics of state math anxiety: (1) anticipation-related mechanisms, which initially drive anxiety, but become reduced as tasks are successfully performed, and (2) habituation-related mechanisms, which foster a gradual decrease in anxiety over time. Broader anxiety theory, consistent with research optimizing exposure therapy[55], emphasizes that therapeutic gains are significantly enhanced by actively violating negative expectancies, rather than relying solely on repetition. This theoretical perspective raises interesting questions about the relative contributions of expectancy violation versus habituation in math anxiety reduction. By distinguishing these two contributing processes, our results provide new insights into how state math anxiety evolves during mathematics over time and upon task completion.

From a theoretical standpoint, the effectiveness of exposure therapy as an intervention for math anxiety might be enhanced if it incorporates strategies beyond repeated exposure to mathematical tasks[56]. The reason is that it would fail to address the critical role of anticipation-related mechanisms of anxiety. Comprehensive interventions should therefore incorporate strategies to manage pre-task expectation anxiety, such as cognitive restructuring or relaxation techniques, alongside a gradual exposure to and training in mathematics[57]. However, the relative effectiveness of different intervention components such as reducing anticipatory anxiety or facilitating habituation should be empirically tested in future studies, as research on therapeutical approaches for anxiety goes beyond the scope of the current study.

In summary, our findings underscore that math anxiety is situation-dependent (state) as well as a personal characteristic (trait), which interact over time. In individuals with high trait math anxiety, the expectation anxiety is elevated and decreases stepwise while performing mathematics. Besides, our data support the use of the state math anxiety scale[24] as an effective tool for detecting situational differences in anxiety levels towards mathematics. Notably, similar temporal patterns were observed for state anxiety (Fig. S1), suggesting that the decrease in anxiety over time may reflect both math-specific and broader anxiety regulation processes.

The current study replicated the relationship between trait math anxiety and arithmetic performance. Consistent with the processing efficiency theory and the attentional control theory[13,14], higher levels of anxiety were associated with slower response times[9,58,59]. Interestingly, trait math anxiety was only significantly related to response time, not accuracy, and the anxiety-complexity effect[8] was also observed only in response time instead of accuracy (Table S2). In absence of a speed-accuracy trade-off (Table S2), this indicates that math-anxious individuals may respond more slowly without sacrificing accuracy in tasks that are in principle solvable and not too difficult (relatively high accuracy leading to ceiling effects), suggesting that anxiety is primarily associated with disruption in cognitive processing speed rather than performance quality in manageable math tasks. While this is in line with the processing efficiency theory[13,14], stating that anxiety impairs processing efficiency (response time) but not performance effectiveness (response accuracy), other studies found the opposite pattern[60,61]. Together, these results challenge the traditional notion of a speed-accuracy trade-off in anxiety-related tasks[42] and suggest that the relationship between math anxiety, response accuracy, and processing efficiency may vary depending on task conditions and measurement approaches.

The anxiety-complexity effect was also replicated in the current study (Table S2)[4,7,8], with individuals with higher trait math anxiety needing particularly more time to solve complex arithmetic (involving carry or borrow operations). Therefore, math anxiety is disproportionately associated with impaired performance under more cognitively challenging conditions, because the carry and borrow operations require working memory[11] and working memory is limited in math-anxious individuals due to intrusive thoughts[62]. These findings suggest that while math-anxious individuals may maintain accuracy through compensatory strategies, the cognitive load required for emotional regulation can extend processing time and reduce problem-solving depth in complex tasks. This supports the disruption account, which postulates that intrusive thoughts and rumination consume working memory resources, thereby impairing performance, particularly on tasks requiring high cognitive demands[9]. In turn, lower performance may induce higher state math anxiety, potentially contributing to higher trait math anxiety over time. However, this assumption requires further investigation, as state math anxiety was not assessed at the trial level in the current study.

Interestingly, trait math anxiety interacted with task difficulty (Fig. S2), but not with paradigm regarding performance. This difference highlights that trait math anxiety relates to performance in a way that is distinct from its relationship with state math anxiety. Trait math anxiety was associated with math performance dependent on task difficulty, reflecting its persistent relationship with individuals' ability to manage increasingly complex cognitive demands. On the other hand, trait math anxiety explained the paradigm-dependent increase in state math anxiety but not the paradigm-dependent drop in performance. This suggests that trait math anxiety is associated with greater performance challenges posed by task complexity but not by task format. It should be noted that other individual differences beyond trait math anxiety, such as test anxiety, math ability, math self-concept, working memory capacity, and gender may also moderate these effects. Our relatively homogeneous and high-performing sample may have limited variability in such factors, reflecting a need for future studies.

In conclusion, trait math anxiety was associated with lower performance, especially in more complex arithmetic. Thus, the anxiety-complexity effect was replicated, with math anxiety being associated with reduced processing efficiency without a decrease in performance effectiveness.

This study offers important educational implications, emphasizing the role of task design in managing math anxiety. Integrating multiple-choice or game-based learning approaches, which have proven effective in typically developing and dyscalculic children[60,63], can reduce cognitive load and anxiety, particularly for those with high trait math anxiety. Additionally, a repeated exposure to mathematical tasks can help reduce anxiety over time, as initial anxiety is often higher before students are familiar with the type of task they will be completing.

Future research might further investigate the mechanisms behind the interaction between math anxiety and task complexity, as well as the temporal dynamics of anxiety during more extended or varied mathematical tasks. Experimentally manipulating anticipatory anxiety would provide more direct causal evidence for the mechanisms we observed. For instance, varying pre-task instructions to increase or decrease performance expectations[44], implementing anxiety induction procedures[64], or testing brief anxiety-reduction interventions[65] could help isolate the specific contribution of anticipatory mechanisms to math anxiety dynamics. Such experimental approaches would complement our correlational findings and provide stronger evidence for designing targeted interventions.

Our study is limited by several points. Regarding task complexity, the two-digit arithmetic tasks used in this study may not have been sufficiently challenging for adults, potentially contributing to the observed decrease in state anxiety as participants adapted to the task and perceived it as less difficult than expected. This habituation effect might differ if more complex math tasks were used (such as those involving fractions, larger numbers or advanced mathematics), or if time constraints were introduced (that might induce stress). Especially the task-phase decline of state math anxiety (pre-, mid-, and post-task) may be specific to tasks that are mastered well (high accuracy). Otherwise, if tasks would be more or too difficult, expectation anxiety at the beginning of the task may not decrease as much or may even increase when individuals fail most of the time. Notably, the high accuracy across paradigms indicates a potential ceiling effect, which may have reduced variability in performance and limited sensitivity to detect subtle relations with anxiety measures. Future studies could therefore consider adopting more demanding tasks to better capture individual differences in math performance.

For sample diversity, the relatively small and homogenous adult sample may limit the variance in our study to interindividual differences and the generalizability of the results, particularly to developmental stages. Specifically, our university student sample showed relatively high math competence (as evidenced by high accuracy rate), limiting variability in math ability. Additionally, the unbalanced gender distribution (71% female) should be noted, as females typically report higher math anxiety[66]. We also did not assess other potentially relevant individual differences such as working memory capacity. Future research should include a more diverse population in terms of age, educational background, and cultural context, as well as standardized measures of math ability to disentangle the effects of math ability from math anxiety, to validate the findings across different groups. Neurocognitive approaches would facilitate the development of more effective, targeted interventions to reduce math anxiety and improve learning outcomes for both typically developing and math-disabled students[63].

As for feedback mechanisms, this study did not incorporate feedback for each arithmetic problem. Research shows that feedback can have complex effects on emotions and learning outcomes, while some studies indicate that feedback, especially corrective feedback, can evoke negative emotions and influence learning outcomes[67], other research suggests that feedback may reduce math anxiety, particularly in higher education settings[68]. Incorporating different types of feedback in future studies would provide a deeper understanding of how feedback impacts state math anxiety and performance.

With respect to self-reported measures, relying solely on self-reported measures of math anxiety may introduce bias. For instance, the fact that women report more anxiety in self-reports than men have been attributed to a response bias with men feeling more uncomfortable to admit anxiety[69]. Moreover, there is a discrepancy between state and trait math anxiety with girls reporting more trait but not state math anxiety than boys[34], and autistic boys reporting more state but not trait math anxiety than non-autistic boys[70]. Future research should adopt multidimensional assessments, combining behavioral experiments and self-report questionnaires with psychophysiological markers and neuroimaging data to obtain more objective and comprehensive data[24,71].

To conclude, this study reveals the dynamic nature of state math anxiety in arithmetic tasks, focusing on the dependency on trait math anxiety, the effects of paradigms, and temporal trends. The results show that production paradigms, which require generating answers, increase state math anxiety compared to decision paradigms, especially in individuals with higher trait math anxiety. Over time, state math anxiety decreases linearly from pre- to post-arithmetic tasks and during repetitions of the task. The anxiety-complexity effect was observed, with higher trait math anxiety being associated with worse performance particularly in complex arithmetic. These findings offer valuable insights for designing educational approaches that reduce anxiety and enhance learning outcomes.

Methods

Participants

The sample included 65 adults (17 male, 46 female, 2 diverse; age: $M = 22.86$ years, $SD = 3.80$ years, $Range = 19–37$ years), taken from the study on paradigms[40]. Among all subjects, 57 were right-handed and 8 were left-handed. Inclusion criteria for participants were an age between 18 and 40 years, native German speakers, and no dyscalculia or other learning disorders (e.g., attention deficit hyperactivity disorder). For participation, all subjects received student credits or monetary reimbursement. Informed written consent was obtained from all subjects and the study was conducted following the latest version of the Declaration of Helsinki.

Material

To assess trait and state (math) anxiety, we administered the following questionnaires. The Abbreviated Math Anxiety Scale (AMAS)[72] was used to measure trait math anxiety. The questionnaire consists of 9 items with a 5-point Likert scale, ranging from 1 (low anxiety) to 5 (high anxiety). The scale demonstrated strong internal consistency (Cronbach's $\alpha = 0.90$), good test-retest reliability ($r = 0.85$), and good convergent and divergent validity[72]. In the present study, a German translated version was used, which also showed very good internal reliability (Cronbach's $\alpha = 0.89$ and ordinal $\alpha = 0.94$).

The State Math Anxiety Scale (SMA)[24] was used to measure state math anxiety. The questionnaire consists of 7 items that capture the emotional, cognitive, and physiological aspects of math anxiety and 2 control items for enjoyment and boredom, with a 5-point Likert scale ranging from 1 (not at all) to 5 (very much). The German scale demonstrated strong internal consistency (Cronbach's $\alpha > .90$, ranging from 0.91 to 0.95) and good validity[24]. As the SMA is situational with questions relating to the math task at hand, some questions would be inappropriate if not presented during the context of a math task (before and after the math task). Therefore, we only kept one item of the SMA (SMA-1) to measure state math anxiety before and after the math task: "How math-anxious do you feel right now?" (adapted from SIMA[73]). In the present study, the SMA also showed good internal consistency (Cronbach's $\alpha \geq 0.80$ during the arithmetic task, with ordinal α ranging from 0.70 to 0.90).

The short German version of the State-Trait Anxiety Inventory (STAI-SKD)[74] was used to measure state anxiety. The questionnaire consists of 5 items reflecting the current emotional state, which should be rated on a 5-point Likert scale ranging from 1 (not at all) to 4 (very much). The internal consistency of the STAI-SKD was satisfactory (Cronbach's $\alpha = 0.76$)[74]. In the present study, the STAI-SKD also demonstrated good internal consistency (Cronbach's α and ordinal $\alpha \geq 0.80$ at each measurement time point). Both state math anxiety and state anxiety were measured to distinguish math-specific emotional responses from broader anxiety dynamics during task performance; however, the constructs of state math anxiety and state anxiety in the math context can be considered the same[24].

All questionnaires were administered using a paper-and-pencil format.

The arithmetic task followed the same procedure as described in our previous study[24]. Each arithmetic problem consisted of two two-digit operands that resulted in a two-digit solution. In a 2×2 design, the problems included addition with (e.g., $36 + 27$) or without (e.g., $32 + 24$) carrying (a carry operation in addition is required when the sum of the units of the operands exceeds 9, with a decade to be carried over) and subtraction with (e.g., 63–25) or without (e.g., 69–23) borrowing (a borrow operation in subtraction is required whenever the unit of the subtrahend is larger than the unit of the minuend, and hence a decade has to be borrowed). Each of these four conditions consisted of 24 arithmetic problems, resulting in 96 problems within one stimulus set. Six stimulus sets were created[24] and were matched in the numerical magnitude of the operands and the result, overall problem size, and counterbalanced in the position of the larger operand. The stimulus sets did not entail trivial cases such as pure decades (e.g., 20), ties (e.g., 22), or unit/decade repetitions. Subtraction problems were constructed as inverse addition problems. Note that the stimulus sets were matched, but not identical to avoid trial-specific learning across paradigms. The distractor in decision paradigms differed from the target at the unit position (± 2) or at the decade position (± 10). The dependent variables for arithmetic performance are accuracy (ACC) and response time (RT).

Two types of paradigms were employed in this study: decision vs. production paradigms (see Fig. 3). In decision paradigms (verification, forced-choice, delayed forced-choice), participants should decide whether the given answer was correct or select the correct answer. In production paradigms (written production, verbal-keyboard production, simple verbal production), participants should calculate the answer and type it or say it aloud. The six specific paradigms used were the following (for a detailed description of the paradigms see Yao et al.[24]): In the verification paradigm, participants need to indicate whether the given answer is right or wrong. In the forced-choice paradigm, participants need to choose one out of two given answer options presented simultaneously with the arithmetic problem. In the delayed forced-choice paradigm, first the arithmetic problem was shown until participants pressed the space bar indicating that they had calculated the answer in mind; afterwards, the two answer options (target and distractor) appeared from which participants chose the calculated answer (with a time limit of 2000 ms). In the written production paradigm, participants need to type the answer directly in a number keyboard. In the verbal-keyboard production paradigm, participants need to verbalize the answer while pressing a button on the keyboard (to record response time). In the simple verbal production paradigm, participants need to verbalize the answer directly (while response time is recorded by voice key).

Procedure

The procedure started with questionnaires on trait math anxiety (AMAS), state math anxiety (SMA-1), and state anxiety (STAI-SKD). Then the arithmetic task was conducted computer-based using the OpenSesame 3.3.10[76]. Each of the 3 decision and 3 production paradigms were presented in alternating blocks (see Fig. 4), with the order of paradigm blocks following a Latin Square design[40], i.e., participants were evenly assigned to six counterbalanced sequences in which each paradigm appears once in each ordinal position. This controls for position effects, though it does not completely counterbalance all possible orders. Each participant completed all six paradigms (one stimulus set of 96 problems per paradigm), resulting in a total of 576 arithmetic problems per participant. The arithmetic problems within each block were presented in randomized order. The participants were instructed to solve the arithmetic problems as quickly and accurately as possible. Within each paradigm block, there was a break after half of the arithmetic problems, in which participants were asked to fill in the questionnaires on state math anxiety (SMA) and state anxiety (STAI-SKD), resulting in 6 measurements during the arithmetic task. After the arithmetic task, state math anxiety (SMA-1) and state anxiety (STAI-SKD) were assessed again (for a flowchart of the experiment see Fig. 4). The whole experiment lasted ~2 h.

Data analysis

Data exclusion criteria were preregistered and applied as follows (for details see Yao et al.[75], Table S1): Participants were case-wise removed with missing data in the arithmetic task, with an ACC below 50% per production paradigm or below 75% per decision paradigm (due to a 50% chance level), or with a mean RT >3 median absolute

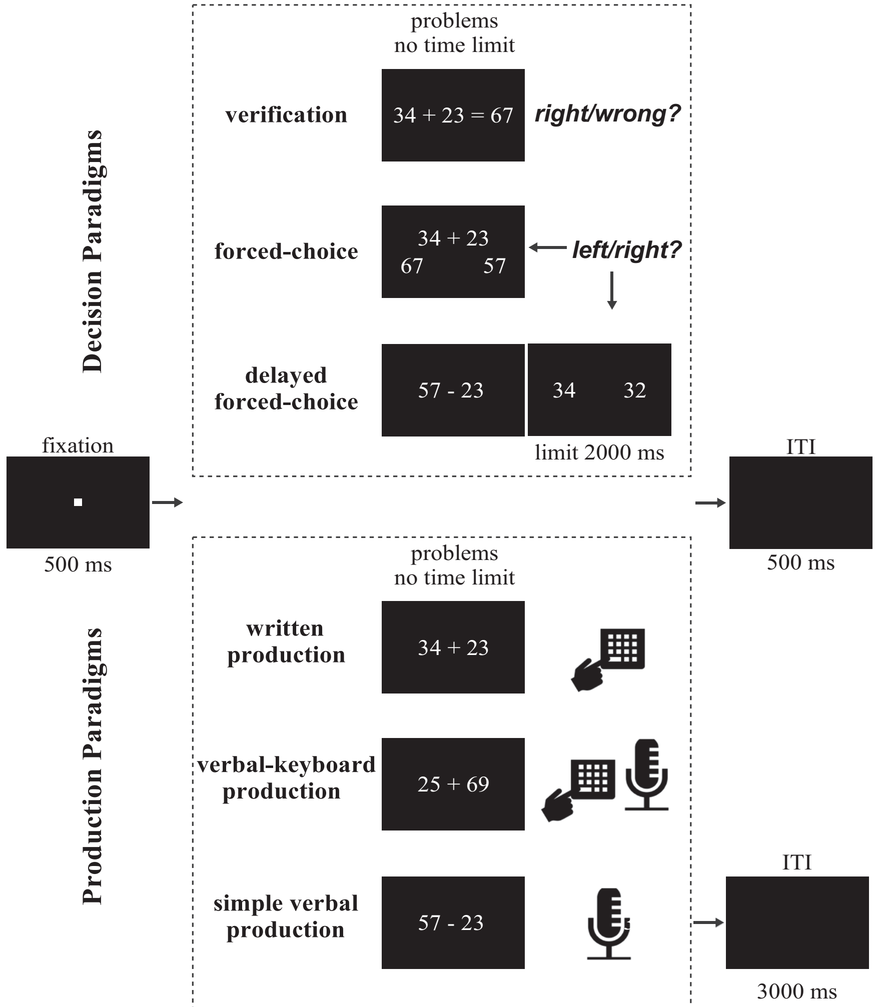


Fig. 3 | Different arithmetic paradigms. Each trial began with a 500 ms fixation, followed by the presentation of an arithmetic problem. In decision paradigms (verification, forced-choice, delayed forced-choice), participants judged the correctness of a presented solution or selected the correct answer from two alternatives. The delayed forced-choice paradigm additionally required participants to indicate by pressing the spacebar on the keyboard when they had mentally calculated the answer before response options appeared (time limit was 2000 ms). In production paradigms (written production, verbal-keyboard production, simple verbal production), participants generated the answer either by typing or verbalizing it. An inter-trial interval (ITI) of 500 ms was used for all paradigms except the simple verbal production paradigm, which used 3000 ms. No time limit was imposed for problem solving, except for the response in the delayed forced-choice condition.

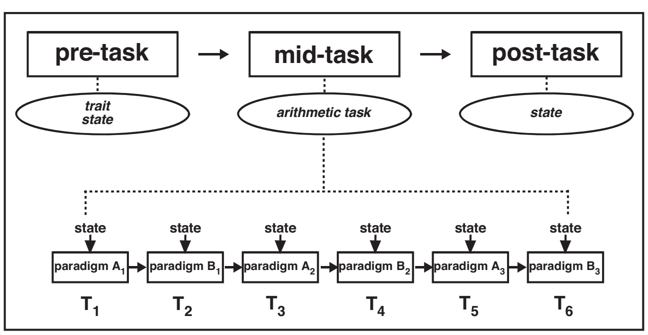


Fig. 4 | Flowchart of the study procedure. T_1 to T_6 indicate the six measurement time points during the arithmetic task. Trait math anxiety was measured pre-task, state math anxiety with one item and state anxiety were measured pre-task and post-task. State (math) anxiety was measured during the break of each paradigm after half of the arithmetic problems. Paradigm A and B refer to the two paradigm types (production vs. decision) presented in alternating order. Subcategories A1, A2, A3 and B1, B2, B3 further distinguish the three production paradigms and the three decision paradigms. The alternating order of paradigms was counterbalanced across participants.

deviation (*MAD*)[76] above or below the group *Median* for the respective paradigm. According to these criteria, the final sample size ranged from 59−65 participants across paradigms (verification: $N = 59$; forced-choice: $N = 59$; delayed forced-choice: $N = 63$; written production: $N = 63$; verbal-keyboard production: $N = 65$; simple verbal production: $N = 64$). In the arithmetic task, trials were removed from RT analysis according to the following criteria: false equation trials in the verification paradigm (also for ACC), incorrectly solved trials (i.e., errors and missing; 8%), RTs below 200 ms (anticipations; 0%), RTs >3 *MAD* above or below the individual *Median* for the respective paradigm (outliers; 6%), or distance between first and second RT >3 *MAD* above or below the individual *Median* for the respective paradigm (delayed forced-choice, written production, and verbal-keyboard production; 0%).

Trial-level accuracy data (0/1) were analyzed using generalized linear mixed-effects models (GLMMs) with a binomial distribution and logit link function[77]. For analyses at the condition- or subject-level (speed–accuracy LMM analysis and exploratory correlation analysis in the current study), logit-transformed accuracy data were used to stabilize variance and approximate a normal distribution. No transformation was applied to RT data, as their distributions were sufficiently normalized following the MAD-based outlier removal procedure.

In the questionnaires, missing data were handled by calculating mean instead of sum scores, with higher values indicating higher anxiety. Missing data were handled by calculating mean instead of sum scores, with higher values indicating higher anxiety. No more than one item was missing per questionnaire and participant. For data analysis, we centered continuous predictors, including trait math anxiety (AMAS), response time (for speed-accuracy trade-off analysis), and time (3 time points and 6 time points) data to facilitate the interpretation of their effects on the dependent variable and to improve the stability of the statistical model[78]. To center the variables, we subtracted the mean value of the variable in our sample from the mean value of each participant so that the centered score represents the deviation of the individual value from the overall sample.

The preregistered data analysis was conducted using the statistical computing software R[79], including the R packages tidyr[80], dplyr[81], ggplot2[82], afex[83], and lmerTest[84]. Given the hierarchical nature of the data, with multiple observations nested within participants, we used Linear Mixed Models (LMMs) and Generalized Linear Mixed Models (GLMMs) to analyze the data. This approach is widely established in cognitive and arithmetic research because it accounts for both between- and within-participant variability and avoids the independence assumptions inherent in traditional ANOVAs[85,86]. All Linear Mixed Models (LMM) and Generalized Linear Mixed Models (GLMM) were fitted with the function lmer from the lmerTest R package, with maximum-likelihood estimation for the fixed effects and logit as link function for the GLMM. As preregistered, we conducted model selection using a top-down strategy[87] based on the ANOVA, which provides a principled way to balance model fit and parsimony. We first fitted a full model, then sequentially reduced the random effect structure, followed by reducing the fixed effect structure, and finally reported the final model. Detailed model specifications and all selection steps are reported in Table S3.

Data availability

The materials, data, and analysis code (R scripts) for this study are openly available on the Open Science Framework (OSF) at https://osf.io/z8cqm/overview.

Code availability

The R code for analysis for each hypothesis are available in the OSF repository (https://osf.io/z8cqm/overview).

Received: 10 July 2025; Accepted: 26 December 2025;
Published online: 20 January 2026

References

1. Cresswell, C. & Speelman, C. P. Does mathematics training lead to better logical thinking and reasoning? a cross-sectional assessment from students to professors. *PLoS ONE* **15**, e0236153 (2020).
2. Jacinto, H. & Carreira, S. Knowledge for teaching mathematical problem-solving with technology: an exploratory study of a mathematics teacher's proficiency. *Eur. J. Sci. Math. Educ.* **11**, 105–122 (2023).
3. Parsons, S. & Bynner, J. Numeracy and employment. *Educ. Train.* **39**, 43–51 (1997).
4. Ashcraft, M. H. & Faust, M. W. Mathematics anxiety and mental arithmetic performance: an exploratory investigation. *Cogn. Emot.* **8**, 97–125 (1994).
5. Barroso, C. et al. A meta-analysis of the relation between math anxiety and math achievement. *Psychol. Bull.* **147**, 134–168 (2021).
6. Brunner, M., Preckel, F., Götz, T., Lüdtke, O. & Keller, L. *The Relationship Between Math Anxiety and Math Achievement: New Perspectives From Combining Individual Participant Data and Aggregated Data in a Meta-Analysis.* http://osf.io/preprints/psyarxiv/6pmhr_v2 (2023).
7. Ashcraft, M. H. & Krause, J. A. Working memory, math performance, and math anxiety. *Psychon. Bull. Rev.* **14**, 243–248 (2007).
8. Huber, J. F. & Artemenko, C. Anxiety-related difficulties with complex arithmetic. *Z. Für Psychol*. **229**, 236–240 (2021).
9. Ashcraft, M. H. & Kirk, E. P. The relationships among working memory, math anxiety, and performance. *J. Exp. Psychol. Gen.* **130**, 224 (2001).
10. Ramirez, G., Shaw, S. T. & Maloney, E. A. Math anxiety: past research, promising interventions, and a new interpretation framework. *Educ. Psychol.* **53**, 145–164 (2018).
11. Imbo, I., Vandierendonck, A. & Vergauwe, E. The role of working memory in carrying and borrowing. *Psychol. Res.* **71**, 467–483 (2007b).
12. Korem, N., Cohen, L. D. & Rubinsten, O. The link between math anxiety and performance does not depend on working memory: a network analysis study. *Conscious. Cogn.* **100**, 103298 (2022).
13. Eysenck, M. W. & Calvo, M. G. Anxiety and performance: the processing efficiency theory. *Cogn. Emot.* **6**, 409–434 (1992).
14. Eysenck, M. W., Derakshan, N., Santos, R. & Calvo, M. G. Anxiety and cognitive performance: attentional control theory. *Emotion* **7**, 336–353 (2007).
15. Ferguson, A. M., Maloney, E. A., Fugelsang, J. & Risko, E. F. On the relation between math and spatial ability: the case of math anxiety. *Learn. Individ. Differ.* **39**, 1–12 (2015).
16. Hembree, R. The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* **21**, 33–46 (1990).
17. Maloney, E. A. Math Anxiety: Causes, Consequences, and Remediation. in *Handbook of Motivation at School* (eds. Wentzel, K. R. & Miele, D. B.) 408−423 (Routledge, 2016).
18. Devine, A., Hill, F., Carey, E. & Szűcs, D. Cognitive and emotional math problems largely dissociate: pevalence of developmental dyscalculia and mathematics anxiety. *J. Educ. Psychol.* **110**, 431–444 (2018).
19. Carey, E., Hill, F., Devine, A. & Szücs, D. The Chicken or the Egg? The Direction of the Relationship Between Mathematics Anxiety and Mathematics Performance. *Front. Psychol*. https://doi.org/10.3389/fpsyg.2015.01987 (2016).
20. Cipora, K., Santos, F. H., Kucian, K. & Dowker, A. Mathematics anxiety-where are we and where shall we go? *Ann. N. Y. Acad. Sci.* **1513**, 10–20 (2022).
21. Spielberger, C. D. *Anxiety: Current Trends in Theory and Research*. (Elsevier, 2013).
22. Orbach, L., Herzog, M. & Fritz, A. Relation of state-and trait-math anxiety to intelligence, math achievement and learning motivation. *J. Numer. Cogn.* **5**, 371–399 (2019).
23. Orbach, L., Herzog, M. & Fritz, A. State-and trait-math anxiety and their relation to math performance in children: the role of core executive functions. *Cognition* **200**, 104271 (2020).
24. Artemenko, C., Cipora, K. & Nuerk, H.-C. *Math Anxiety Increases With The Difficulty Of A Math Task – The State-trait Differentiation For Math Anxiety*. https://osf.io/preprints/osf/f269d_v2 (2025).
25. Robinson, M. D. & Clore, G. L. Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychol. Bull.* **128**, 934 (2002).
26. Roos, A.-L. et al. Experiencing more mathematics anxiety than expected? contrasting trait and state anxiety in high achieving students. *High Abil. Stud.* **26**, 245–258 (2015).
27. Trezise, K. & Reeve, R. A. Patterns of anxiety in algebraic problem solving: a three-step latent variable analysis. *Learn. Individ. Differ.* **66**, 78–91 (2018).
28. Punaro, L. & Reeve, R. Relationships between 9 year-olds′ math and literacy worries and academic abilities. *Child Dev. Res.* **2012**, 359089 (2012).
29. Daches Cohen, L., Korem, N. & Rubinsten, O. Math anxiety is related to math difficulties and composed of emotion regulation and anxiety predisposition: a network analysis study. *Brain Sci* **11**, 1609 (2021).
30. Strohmaier, A. R., Schiepe-Tiska, A. & Reiss, K. M. A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. An application of the control-value theory. *Frontline Learn. Res.* **8**, 16–32 (2020).
31. Daker, R. J., Gattas, S. U., Necka, E. A., Green, A. E. & Lyons, I. M. Does anxiety explain why math-anxious people underperform in math? *Npj Sci. Learn.* **8**, 6 (2023).
32. Pelegrina, S., Martín-Puga, M. E., Lechuga, M. T., Justicia-Galiano, M. J. & Linares, R. Role of executive functions in the relations of state- and trait-math anxiety with math performance. *Ann. N. Y. Acad. Sci.* **1535**, 76–91 (2024).
33. Bieg, M., Goetz, T. & Lipnevich, A. A. What students think they feel differs from what they really feel–academic self-concept moderates

the discrepancy between students' trait and state emotional self-reports. *Plos One* **9**, e92563 (2014).
34. Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R. & Hall, N. C. Do girls really experience more anxiety in mathematics? *Psychol. Sci.* **24**, 2079–2087 (2013).
35. Schmidt, P., Jendryczko, D., Zurbriggen, C. L. A. & Nussbeck, F. W. Recall bias of students' affective experiences in adolescence: The role of personality and internalizing behavior. *J. Adolesc.* **95**, 893–906 (2023).
36. Bieg, M., Goetz, T., Wolter, I. & Hall, N. C. Gender stereotype endorsement differentially predicts girls' and boys' trait-state discrepancy in math anxiety. *Front. Psychol*. https://doi.org/10.3389/fpsyg.2015.01404 (2015).
37. Caviola, S. et al. Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: a meta-analysis on 906,311 participants. *Educ. Psychol. Rev.* **34**, 363–399 (2022).
38. Malik, U. Effect of multiple choice question format on student performance. *Proc. Aust. Conf. Sci. Math. Educ.* **123**, 128 (2020).
39. Zhang, J., Zhao, N. & Kong, Q. P. The relationship between math anxiety and math performance: a meta-analytic investigation. *Front. Psychol.* **10**, 458192 (2019).
40. Yao, X., Artemenko, C., He, Y. & Nuerk, H.-C. Arithmetic is not arithmetic: paradigm matters for arithmetic effects. *Cognition* **256**, 106060 (2025).
41. Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U. & Nieuwenhuis, S. The neural basis of the speed–accuracy tradeoff. *Trends Neurosci.* **33**, 10–16 (2010).
42. Wickelgren, W. A. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol. (Amst.)* **41**, 67–85 (1977).
43. Orbach, L. & Fritz, A. Patterns of attention and anxiety in predicting arithmetic fluency among school-aged children. *Brain Sci.* **12**, 376 (2022).
44. Lyons, I. M. & Beilock, S. L. Mathematics anxiety: separating the math from the anxiety. *Cereb. Cortex* **22**, 2102–2110 (2012).
45. Lyons, I. M. & Beilock, S. L. When math hurts: math anxiety predicts pain network activation in anticipation of doing math. *PLoS ONE* **7**, e48076 (2012).
46. Goetz, T., Preckel, F., Pekrun, R. & Hall, N. C. Emotional experiences during test taking: does cognitive ability make a difference? *Learn. Individ. Differ.* **17**, 3–16 (2007).
47. Conlon, R. A., Hicks, A., Barroso, C. & Ganley, C. M. The effect of the timing of math anxiety measurement on math outcomes. *Learn. Individ. Differ.* **86**, 101962 (2021).
48. Qu, Z. et al. Measurement of high-school students' trait math anxiety using neurophysiological recordings during math exam. *IEEE Access* **8**, 57460–57471 (2020).
49. Benito, K. G. & Walther, M. Therapeutic process during exposure: Habituation model. *J. Obsessive-Compuls. Relat. Disord.* **6**, 147–157 (2015).
50. Endler, N. S. & Kocovski, N. L. State and trait anxiety revisited. *J. Anxiety Disord.* **15**, 231–245 (2001).
51. Koponen, T. K. et al. Does multi-component strategy training improve calculation fluency among poor performing elementary school children? *Front. Psychol*. https://doi.org/10.3389/fpsyg.2018.01187 (2018).
52. Koponen, T. et al. Benefits of integrating an explicit self-efficacy intervention with calculation strategy training for low-performing elementary students. *Front. Psychol*. https://doi.org/10.3389/fpsyg.2021.714379 (2021).
53. Ramirez, G., Gunderson, E. A., Levine, S. C. & Beilock, S. L. Math anxiety, working memory, and math achievement in early elementary school. *J. Cogn. Dev.* **14**, 187–202 (2013).
54. Endler, N. S. Stress, anxiety and coping: the multidimensional interaction model. *Can. Psychol* **38**, 136–153 (1997).
55. Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T. & Vervliet, B. Maximizing exposure therapy: an inhibitory learning approach. *Behav. Res. Ther.* **58**, 10–23 (2014).
56. Passolunghi, M. C., De Vita, C. & Pellizzoni, S. Math anxiety and math achievement: the effects of emotional and math strategy training. *Dev. Sci.* **23**, e12964 (2020).
57. Sammallahti, E., Finell, J., Jonsson, B. & Korhonen, J. A meta-analysis of math anxiety interventions. *J. Numer. Cogn.* **9**, 346–362 (2023).
58. Chang, H. & Beilock, S. L. The math anxiety-math performance link and its relation to individual and environmental factors: a review of current behavioral and psychophysiological research. *Curr. Opin. Behav. Sci.* **10**, 33–38 (2016).
59. Demedts, F., Reynvoet, B., Sasanguie, D. & Depaepe, F. Unraveling the role of math anxiety in students' math performance. *Front. Psychol*. https://doi.org/10.3389/fpsyg.2022.979113 (2022).
60. Soltanlou, M. et al. Math anxiety in combination with low visuospatial memory impairs math learning in children. *Front. Psychol*. https://doi.org/10.3389/fpsyg.2019.00089 (2019).
61. Vukovic, R. K., Kieffer, M. J., Bailey, S. P. & Harari, R. R. Mathematics anxiety in young children: concurrent and longitudinal associations with mathematical performance. *Contemp. Educ. Psychol.* **38**, 1–10 (2013).
62. Ashcraft, M. H. Math anxiety: personal, educational, and cognitive consequences. *Curr. Dir. Psychol. Sci.* **11**, 181–185 (2002).
63. Soltanlou, M. et al. Training causes activation increase in temporo-parietal and parietal regions in children with mathematical disabilities. *Brain Struct. Funct.* **227**, 1757–1771 (2022).
64. Lallement, C. & Lemaire, P. Age-related differences in how negative emotions influence arithmetic performance. *Cogn. Emot.* **35**, 1382–1399 (2021).
65. Balt, M., Börnert-Ringleb, M. & Orbach, L. Reducing math anxiety in school children: a systematic review of intervention research. *Front. Educ*. https://doi.org/10.3389/feduc.2022.798516 (2022).
66. Rossi, S. et al. Mathematics–gender stereotype endorsement influences mathematics anxiety, self-concept, and performance differently in men and women. *Ann. N. Y. Acad. Sci.* **1513**, 121–139 (2022).
67. Wagner, S. et al. The more, the better? Learning with feedback and instruction. *Learn. Instr.* **89**, 101844 (2024).
68. Núñez-Peña, M. I., Bono, R. & Suárez-Pellicioni, M. Feedback on students' performance: a possible way of reducing the negative effect of math anxiety in higher education. *Int. J. Educ. Res.* **70**, 80–87 (2015).
69. Sigmon, S. T. et al. Gender differences in self-reports of depression: the response bias hypothesis revisited. *Sex Roles* **53**, 401–411 (2005).
70. Lievore, R. & Mammarella, I. C. Trait and state mathematics anxiety in autistic and non-autistic school-aged boys. *Autism* **29**, 1209−1223 (2024).
71. Mammarella, I. C., Caviola, S., Rossi, S., Patron, E. & Palomba, D. Multidimensional components of (state) mathematics anxiety: Behavioral, cognitive, emotional, and psychophysiological consequences. *Ann. N. Y. Acad. Sci.* **1523**, 91–103 (2023).
72. Hopko, D. R., Mahadevan, R., Bare, R. L. & Hunt, M. K. The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment* **10**, 178–182 (2003).
73. Núñez-Peña, M. I., Guilera, G. & Suárez-Pellicioni, M. The single-item math anxiety scale: an alternative way of measuring mathematical anxiety. *J. Psychoeduc. Assess.* **32**, 306–317 (2014).
74. Englert, C., Bertrams, A. & Dickhäuser, O. Entwicklung Der fünf-item-kurzskala STAI-SKD Zur messung von zustandsangst. *Z. Für Gesundheitspsychologie* **19**, 173–180 (2011).
75. Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* **44**, 314–324 (2012).

76. Bayot, M. et al. The interaction between cognition and motor control: a theoretical framework for dual-task interference effects on posture, gait initiation, gait and turning. *Neurophysiol. Clin. Clin. Neurophysiol.* **48**, 361–375 (2018).
77. Jaeger, T. F. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* **59**, 434–446 (2008).
78. Schielzeth, H. Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**, 103–113 (2010).
79. R Core Team. *R: A Language and Environment For Statistical Computing*. http://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf (2022).
80. Wickham, H., Vaughan, D. & Girlich, M. tidyr. *Tidy Messy Data*, https://tidyr.tidyverse.org/ (2024).
81. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: A Grammar of Data Manipulation. (2023).
82. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis* 2nd edn, Vol. 260 (Springer-Verlag New York, 2016).
83. Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. S. *afex: Analysis of Factorial Experiments*. http://github.com/singmann/afex (2023).
84. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw*. **82**, 13 (2017).
85. Baayen, R. H., Davidson, D. J. & Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* **59**, 390–412 (2008).
86. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
87. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* 1st edn, vol. 574 (Springer, 2009).

Acknowledgements

We would like to thank the participants in this study. We are also grateful to Annalena Wels for her help in data collection. XY, CA, and HCN are members of the LEAD Graduate School & Research Network (GSC1028, funded by the Excellence Initiative of the German federal and state governments). CA was supported by the European Social Fund and the Ministry of Science, Research and the Arts Baden-Wuerttemberg, by the German Research Foundation (DFG, grant number: 468460838, AR 1500/1-1; grant number: 513458453, AR 1500/2-1), and by the Tuebingen Postdoc Academy for Research on Education (PACE) at the Hector Research Institute of Education Sciences and Psychology. We acknowledge support from the Open Access Publication Fund of the University of Tübingen.

Author contributions

Conceptualization: Xinru Yao, Julia F. Huber, Hans-Christoph Nuerk, Christina Artemenko; Methodology: Xinru Yao, Julia F. Huber, Christina Artemenko, Hans-Christoph Nuerk; Software: Xinru Yao, Julia F. Huber, Zixu Li; Validation: Xinru Yao, Christina Artemenko; Formal analysis: Xinru Yao, Zixu Li, Yaren Findik, Julia F. Huber; Investigation: Xinru Yao; Resources: -; Data Curation: Xinru Yao; Writing – original draft: Xinru Yao; Writing – review & editing: Xinru Yao, Julia F. Huber, Zixu Li, Yaren Findik, Hans-Christoph Nuerk, Christina Artemenko; Visualization: Xinru Yao; Supervision: Hans-Christoph Nuerk, Christina Artemenko; Project Administration: Xinru Yao; Funding Acquisition: Hans-Christoph Nuerk.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41539-025-00398-z.

Correspondence and requests for materials should be addressed to Xinru Yao or Hans-Christoph Nuerk.

A2.2 Supplementary Materials for Study 2

These supplementary materials belong to the following publication:

Yao, X., Huber, J. F., Li, Z., Findik, Y., Nuerk, H.-C., & Artemenko, C. (2026). The dynamics of state math anxiety vary by paradigm and timing during arithmetic. *npj* Science of Learning. Advance online publication. https://doi.org/10.1038/s41539-025-00398-z

Supplementary Materials

Title: The dynamics of state math anxiety vary by paradigm and timing during arithmetic

Table S1. Arithmetic performance and mid-task state anxiety across paradigms

	Decision paradigms			**Production paradigms**		
	verification	forced-choice	delayed forced-choice	written production	verbal-keyboard production	simple verbal production
Arithmetic performance						
Accuracy						
addition simple	0.93 (0.08)	0.96 (0.04)	0.95 (0.06)	0.92 (0.07)	0.96 (0.04)	0.92 (0.08)
addition complex	0.94 (0.06)	0.93 (0.08)	0.92 (0.08)	0.90 (0.08)	0.91 (0.08)	0.88 (0.11)
subtraction simple	0.94 (0.07)	0.95 (0.05)	0.94 (0.05)	0.92 (0.06)	0.92 (0.07)	0.90 (0.10)
subtraction complex	0.91 (0.08)	0.92 (0.07)	0.91 (0.08)	0.87 (0.12)	0.88 (0.10)	0.82 (0.15)
Reaction time						
addition simple	3.20 (0.84)	3.04 (0.64)	2.96 (0.75)	3.86 (0.81)	3.86 (0.81)	3.02 (0.80)
addition complex	4.35 (1.01)	3.87 (0.86)	4.21 (1.17)	5.30 (1.32)	5.30 (1.32)	4.27 (1.03)
subtraction simple	4.03 (1.09)	3.64 (0.96)	3.73 (1.01)	4.69 (1.15)	4.69 (1.15)	3.81 (1.00)
subtraction complex	5.02 (1.23)	4.50 (1.02)	4.84 (1.23)	6.04 (1.33)	6.04 (1.33)	5.01 (1.19)
Questionnaires						
mid-task state math anxiety	1.59 (0.72)	1.54 (0.63)	1.58 (0.70)	1.62 (0.73)	1.82 (0.83)	2.19 (1.00)
mid-task state anxiety	1.55 (0.69)	1.41 (0.50)	1.41 (0.61)	1.43 (0.57)	1.78 (0.76)	2.12 (0.95)

Notes. Values are presented as *M* (*SD*) for each decision and production paradigm. State (math) anxiety was measured during the break at the midpoint of each paradigm of the arithmetic task.

Part 1: Additional analysis and results

Table S2. LMM/GLMM results

Predictors	*β*	*CI*	*t / z*	*p*	R^2
Model S1: LMM for three-time points analysis on state anxiety					.62
(intercept)	1.57	1.46 – 1.69	27.20	**< .001**	
time	-0.27	-0.34 – -0.20	-7.46	**< .001**	
trait math anxiety	0.31	0.14 – 0.48	3.64	**< .001**	
time × trait math anxiety	-0.20	-0.31 – -0.10	-3.86	**< .001**	
time^2					
time^2 × trait math anxiety					
Model S2: LMM for six-time points analysis on state anxiety					.71
(intercept)	1.62	1.36 – 1.89	12.08	**< .001**	
time	-0.09	-0.11 – -0.07	-7.59	**< .001**	
time^2	-0.00	-0.02 – 0.01	-0.23	.820	
trait math anxiety	0.18	-0.03 – 0.39	1.67	.096	
time^2 × trait math anxiety	0.03	0.00 – 0.05	2.10	**.036**	
time × trait math anxiety					
Model S3: LMM for trait math anxiety and difficulty on response time					.39
(intercept)	3.38	3.18 – 3.59	32.30	**< .001**	
trait math anxiety	0.30	0.00 – 0.59	1.97	**.049**	
difficulty	1.20	1.13 – 1.28	30.70	**< .001**	
trait math anxiety × difficulty	0.15	0.09 – 0.20	5.46	**< .001**	
Model S4: GLMM for trait math anxiety and difficulty on accuracy					.14
(intercept)	2.83	2.66 – 3.00	32.07	**< .001**	
difficulty	-0.47	-0.57 – -0.38	-9.76	**< .001**	
trait math anxiety					
trait math anxiety × difficulty					
Model S5: GLMM for speed-accuracy trade-off					.68
(intercept)	2.54	2.42 – 2.66	42.31	**< .001**	
paradigm [production]	-0.29	-0.39 – -0.19	-5.77	**< .001**	
trait math anxiety					
response time					
paradigm × trait math anxiety					
paradigm × response time					
trait math anxiety × response time					
paradigm × trait math anxiety × response time					

Notes. *t* value for LMM and *z* value for GLMM. Time and trait math anxiety were centered. Difficulty was dummy coded with simple as reference for complex. Conditional R^2 quantifies the proportion of variance explained by the entire model, including both fixed and random effects.

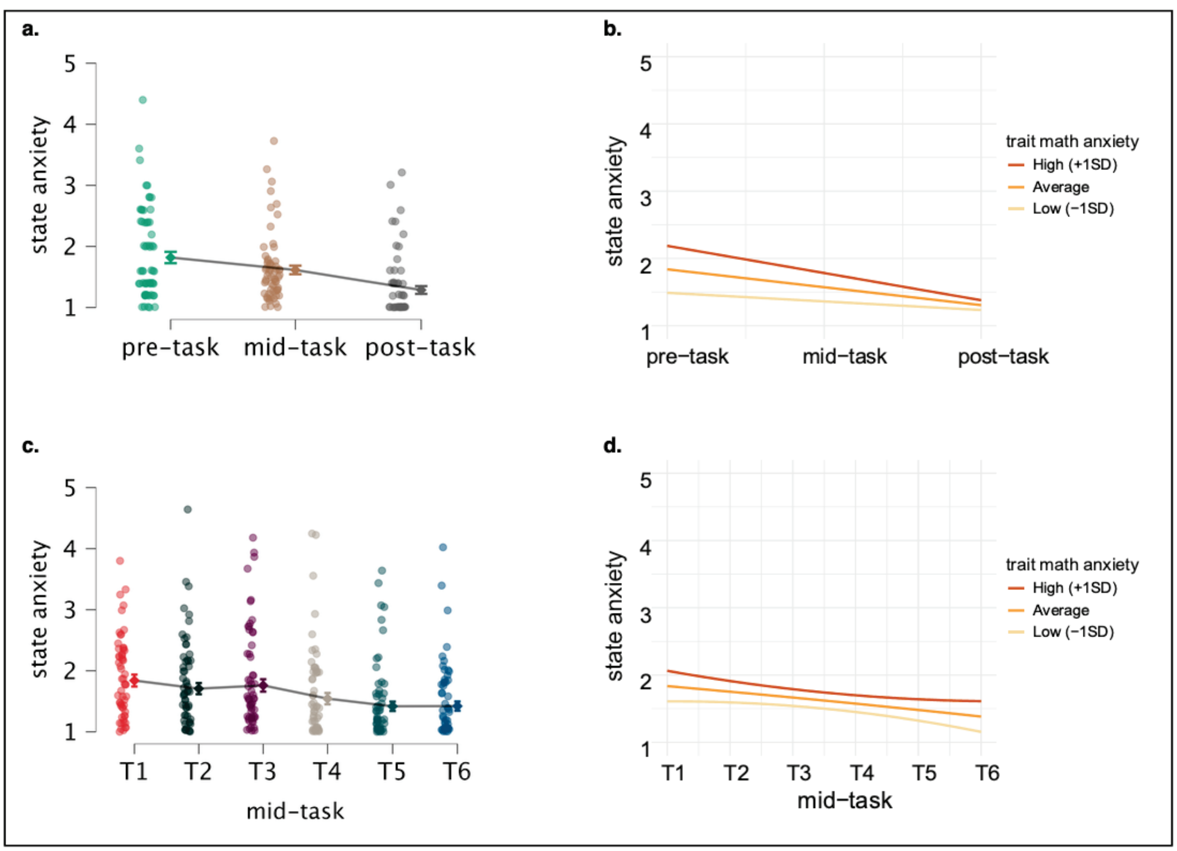
Time analysis for state anxiety pre-, mid-, and post-arithmetic task

Similar to the LMM results for state math anxiety (see Model D in Table 2), the final LMM for state anxiety (see Model S1 in Table S2, Figures S1a & S1b) included fixed effects for time, trait math anxiety, and the interaction between time and trait math anxiety, with a random intercept for subject. The main effect of time indicates that state anxiety decreases over time, with an estimated reduction of -0.27 per time point. The main effect of trait math anxiety shows that for each unit increase in trait math anxiety, state anxiety increases by an estimated 0.31, suggesting that individuals with higher levels of trait math anxiety exhibit higher levels of state anxiety. Additionally, the interaction effect between time and trait math anxiety, with an estimate of -0.20, implies that the decrease in state anxiety over time is more pronounced for individuals with higher trait math anxiety.

Time analysis for state anxiety during the arithmetic task

Different from the LMM results for state math anxiety (see Model E in Table 2), the final LMM for state anxiety (see Model S2 in Table S2, Figures S1c & S1d) included, besides the fixed effect for time, fixed effects for time^2, trait math anxiety, and the interaction between time^2 and trait math anxiety, with random intercepts for subject and paradigm. The main effect of time indicates that state anxiety decreases over time, with an estimated reduction of -0.09 per time point. Different to the larger linear decrease of state math anxiety across six measurement times during arithmetic task for individuals with higher trait math anxiety, the interaction between time^2 and trait math anxiety, with an estimate of 0.03, indicates a non-linear relationship: individuals with higher trait math anxiety show an particularly an initial decrease in state anxiety that stabilizes during the task, reflecting an adaptive or reactive response to the task demands.

Figure S1. State anxiety changes across time



Notes. (a) and (c) show the decrease in state anxiety across the three task phases and across the six measurement times during the arithmetic task, respectively. Error bars represent the standard error of the mean (*SEM*). (b) and (d) show a simple slope analysis for state anxiety depending on trait math anxiety (average level, high level with 1 *SD* above the average, and low level with 1 *SD* below the average) across the three task phases and during the arithmetic task.

Anxiety-complexity effect

An LMM with RT as dependent variable and a GLMM with ACC as dependent variable were conducted including fixed effects for trait math anxiety, difficulty (complex vs. simple), and their interaction. The (G)LMMs further included random intercepts for both subject and item, as well as – not preregistered – a random slope for paradigm, because we found significant differences in performance between paradigms (Yao et al., 2025).

The final LMM for RT (see Model S3 in Table S2, Figure S2) included fixed effects for trait math anxiety, difficulty and their interaction, with random intercepts for subject and item as well as a random slope for paradigm. The main effect of trait math anxiety indicates that for every unit increase in trait math anxiety, the response time increases by an estimate of 0.30 s, so that individuals with higher trait math anxiety take longer to solve arithmetic. The main effect of difficulty indicates that complex arithmetic (with carry/borrow) takes longer to be solved by an estimate of 1.20 s than simple arithmetic (without carry/borrow). The interaction of trait math anxiety and difficulty indicates an anxiety-complexity effect, so that with increasing trait math anxiety the difficulty effect increases by an estimate of 0.15 s.

Different from RT, the final GLMM for ACC (see Model S4 in Table S2) included only a fixed effect for difficulty, with random intercepts for subject and item as well as a random slope for paradigm. The main effect of difficulty indicates that accuracy is higher by an estimate of -0.47 for simple arithmetic (without carry/borrow) compared to complex arithmetic (with carry/borrow).

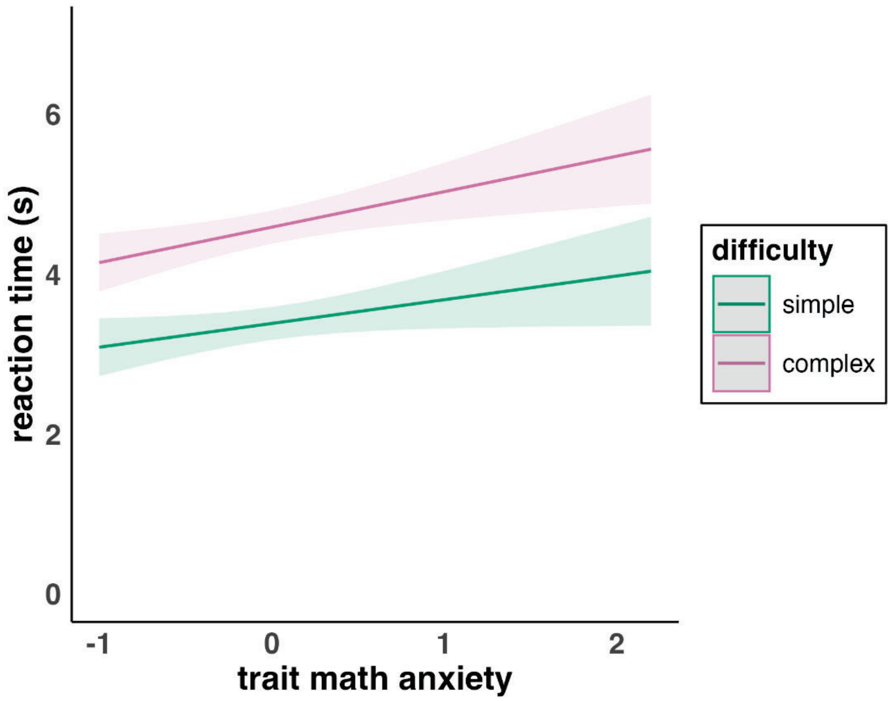
Together, the results suggest that the difficulty effect was replicated in both RT and ACC. For RT, further an anxiety-complexity effect was found, so that individuals with higher trait math anxiety needed longer for arithmetic than individuals with lower math anxiety, especially for more complex arithmetic.

Speed-accuracy trade-off

An LMM with ACC as the dependent variable was conducted (but not a GLMM – as preregistered – because when including response time as a fixed factor, the model failed to converge on a trial level; instead, we performed an LMM on a subject level with average accuracy per participant as dependent variable) including fixed effects for paradigm, RT, trait math anxiety, and their interactions. The LMM further included a random intercept for subject (but not random slopes for RT and paradigm as preregistered, because incorporating this would have made the random effects structure too complex given the limited number of observations).

The final LMM on ACC (see Model S5 in Table S2) only included a fixed effect for paradigm – but not response time. Therefore, the results did not reveal a speed-accuracy trade-off for the current study.

Figure S2. The anxiety-complexity effect



Notes: The anxiety-complexity is indicated by the interaction of trait math anxiety and difficulty on arithmetic performance in terms of response times. Individuals with higher trait math anxiety show higher carry/borrow effects than individuals with lower trait math anxiety, i.e., they need even more time for complex arithmetic (with carry/borrow operation) as compared to simple arithmetic (without carry/borrow operation).

Part 2: Model selection process

Table S3. Full models

Predictors	*β*	*CI*	*t / z*	*p*	R^2
Model A: LMM for paradigm and trait math anxiety on state math anxiety					.87
(intercept)	1.57	1.42 – 1.72	20.74	< **.001**	
paradigm	0.29	0.21 – 0.38	6.64	< **.001**	
trait math anxiety	0.28	0.06 – 0.50	2.52	**.013**	
paradigm × trait math anxiety	0.24	0.11 – 0.37	3.64	< **.001**	
Model B: LMM for trait math anxiety and paradigm on response time					.42
(intercept)	3.96	3.75 – 4.16	37.47	< **.001**	
paradigm	0.41	0.30 – 0.52	7.48	< **.001**	
trait math anxiety	0.36	0.06 – 0.65	2.39	**.017**	
paradigm × trait math anxiety	*0.10*	*0.06 – 0.26*	*1.21*	.226	
Model C: GLMM for trait math anxiety and paradigm on accuracy					.15
(intercept)	3.18	2.74 – 3.62	14.06	< **.001**	
paradigm	-0.58	-0.94 – -0.22	-3.16	**.002**	
trait math anxiety	-0.25	-0.46 – 0.04	-2.37	**.018**	
paradigm × trait math anxiety	*0.15*	*-0.02 – 0.32*	*1.74*	*.081*	
Model D: LMM for three-time points analysis on state math anxiety					.70
(intercept)	1.66	1.49 – 1.83	19.23	< **.001**	
time	-0.31	-0.39 – -0.22	-7.39	< **.001**	
trait math anxiety	0.51	0.26 – 0.77	4.03	< **.001**	
time × trait math anxiety	-0.31	-0.43 – -0.19	-5.10	< **.001**	
$time^2$	*0.11*	*0.03 – 0.25*	*1.52*	*.131*	
$time^2$ × trait math anxiety	*0.13*	*0.08 – 0.34*	*1.24*	*.216*	
Model E: LMM results for six-time points analysis on state math anxiety					.75
(intercept)	1.71	1.47 – 1.96	13.68	< **.001**	
time	-0.11	-0.13 – -0.08	-8.95	< **.001**	
trait math anxiety	0.37	0.14 – 0.60	3.22	**.001**	
time × trait math anxiety	-0.07	-0.11 – -0.04	-3.95	< **.001**	
$time^2$	*0.00*	*0.01 – 0.02*	*0.42*	*.677*	

time^2 × trait math anxiety	*0.01*	*0.01 – 0.03*	*0.79*	*.431*	
Model S1: LMM for three-time points analysis on state anxiety					0.62
(intercept)	1.61	1.47 – 1.75	22.73	**< .001**	
time	-0.27	-0.34 – -0.20	-7.45	**< .001**	
trait math anxiety	0.17	0.18 – 0.51	0.97	.335	
time × trait math anxiety	-0.37	0.75 – 0.00	-1.97	.051	
time^2	*-0.37*	*0.75 – 0.00*	*-1.97*	*.051*	
time^2 × trait math anxiety	*0.09*	*0.10 – 0.27*	*0.93*	*.352*	
Model S2: LMM for six-time points analysis on state anxiety					.71
(intercept)	1.62	1.36 – 1.89	12.08	**< .001**	
time	-0.09	-0.11 – -0.07	-7.60	**< .001**	
time^2	-0.00	-0.02 – 0.01	-0.23	.819	
trait math anxiety	0.18	-0.03 – 0.39	1.67	.096	
time^2 × trait math anxiety	0.03	0.00 – 0.05	2.11	**.036**	
time × trait math anxiety	*-0.02*	*-0.06 – 0.01*	*-1.37*	*.171*	
Model S3: LMM for trait math anxiety and difficulty on response time					.39
(intercept)	3.38	3.18 – 3.59	32.30	**< .001**	
trait math anxiety	0.30	0.00 – 0.59	1.97	**.049**	
difficulty	1.20	1.13 – 1.28	30.70	**< .001**	
trait math anxiety × difficulty	0.15	0.09 – 0.20	5.46	**< .001**	
Model S4: GLMM for trait math anxiety and difficulty on accuracy					.13
(intercept)	17.22	14.57 – 20.35	33.40	**< .001**	
difficulty	0.62	0.57 – 0.69	-9.72	**< .001**	
trait math anxiety	*0.82*	*0.66 – 1.02*	*-1.80*	*.071*	
trait math anxiety × difficulty	*0.97*	*0.87 – 1.08*	*-0.61*	*.542*	
Model S5: GLMM for speed-accuracy trade-off					.67
(intercept)	2.54	2.42 – 2.66	42.31	**< .001**	
paradigm	-0.29	-0.39 – -0.19	-5.77	**< .001**	
trait math anxiety	-0.64	-1.38 – 0.99	-1.73	.086	
response time	-0.02	-0.16 – 0.13	-0.26	.798	
paradigm × trait math anxiety	0.26	-0.51 – 1.04	0.68	.501	
paradigm × response time	-0.07	-0.20 – 0.07	-1.00	.322	

trait math anxiety × response time	0.11	-0.06 – 0.29	1.30	.197
paradigm × trait math anxiety× response time	-0.05	-0.22 – 0.13	-0.53	.595

Notes. *t* value for LMM and *z* value for GLMM for full models. Factors in italics are removed factors compared with reduced final model. Time and trait math anxiety were centered. Difficulty was dummy coded with simple as reference for complex. Conditional R^2 quantifies the proportion of variance explained by the entire model, including both fixed and random effects.

1. Paradigm-dependent analysis for state math anxiety

Full model	*state math anxiety ~ paradigm + trait math anxiety + paradigm : trait math anxiety + (1 | subject)*	
Step 1	*state math anxiety ~ paradigm + trait math anxiety + (1 | subject)*	The model excluding the interaction between paradigm and trait math anxiety was significantly different from the original model (χ^2 = 12.02, $p < .001$). Therefore, we retained the interaction in the final model.
Final Model	**Model A**: *state math anxiety ~ paradigm + trait math anxiety + paradigm : trait math anxiety + (1 | subject)*	Final model = Full model
Random effects	subject (intercept, variance = 0.31, SD = 0.56) (residual, variance = 0.06, SD = 0.25)	

2. Trait math anxiety and paradigm effects on arithmetic performance

(1) LMM for response time:

Full model	*response time ~ trait math anxiety + paradigm + trait math anxiety : paradigm + (1 + paradigm | subject) + (1 | item)*	
Step 1	*response time ~ trait math anxiety + paradigm + trait math anxiety : paradigm + (1 | subject) + (1 | item)*	The model excluding the random slope for paradigm was significantly different from the original model (χ^2 = 334.03, $p < .001$). Therefore, we retained the random slope for paradigm in the final model.
Step 2	*response time ~ trait math anxiety + paradigm + trait math anxiety : paradigm + (1 + paradigm | subject)*	The model excluding the random intercept for item was significantly different from the original model (χ^2 = 4808.20, $p < .001$). Therefore, we retained the random intercept for item in the final model.
Step 3	*response time ~ trait math anxiety + paradigm + (1 + paradigm | subject) + (1 | item)*	The model excluding the interaction of trait math anxiety and paradigm was not significantly different from the original

		model ($\chi^2 = 1.48$, $p = .223$). Therefore, we removed the interaction of trait math anxiety and paradigm in the final model.
Step 4	*response time ~ trait math anxiety + (1 + paradigm | subject) + (1 | item)*	The model excluding the fixed effect of paradigm was significantly different from the original model ($\chi^2 = 42.55$, $p < .001$). Therefore, we retained the fixed effect of paradigm in the final model.
Step 5	*response time ~ paradigm + (1 + paradigm | subject) + (1 | item)*	The model excluding the fixed effect of trait math anxiety was significantly different from the original model ($\chi^2 = 6.92$, $p = .031$). Therefore, we retained the fixed effect of trait math anxiety in the final model.
Final model	**Model B**: *response time ~ trait math anxiety + paradigm + (1 + paradigm | subject) + (1 | item)*	Final model = Step 3
Random effects	item (intercept, variance = 0.66, SD = 0.81) subject (intercept, variance = 0.65, SD = 0.80) paradigm (slope, variance = 0.17, SD = 0.41, $r = 0.07$) (residual, variance = 2.10, SD = 1.45)	

(2) GLMM for accuracy:

Full model	*accuracy ~ trait math anxiety + paradigm + trait math anxiety : paradigm + (1 + paradigm | subject) + (1 | item)*	
Step 1	*accuracy ~ trait math anxiety + paradigm + trait math anxiety : paradigm + (1| subject) + (1 | item)*	The model excluding the random slope for paradigm was significantly different from the original model ($\chi^2 = 29.18$, $p < .001$). Therefore, we retained the random slope for paradigm in the final model.
Step 2	*accuracy ~ trait math anxiety + paradigm + trait math anxiety : paradigm + (1 + paradigm | subject)*	The model excluding the random intercept for item was significantly different from the original model ($\chi^2 = 158.88$, $p < .001$). Therefore, we retained the random intercept for item in the final model.
Step 3	*accuracy ~ trait math anxiety + paradigm + (1 + paradigm | subject) + (1 | item)*	The model excluding the interaction of trait math anxiety and paradigm was not significantly different from the original model ($\chi^2 = 2.89$, $p = .089$). Therefore, we removed the interaction of trait math anxiety and paradigm in the final

		model.
Step 4	*accuracy ~ trait math anxiety + (1 + paradigm | subject) + (1 | item)*	The model excluding the fixed effect of paradigm item was significantly different from the original model (χ^2 = 19.61, p < .001). Therefore, we retained the fixed effect of paradigm in the final model.
Step 5	*accuracy ~ trait math anxiety + paradigm + (1 + paradigm | subject) + (1 | item)*	The model excluding the fixed effect of trait math anxiety was significantly different from the original model (χ^2 = 6.50, p = .039). Therefore, we retained the fixed effect of trait math anxiety in the final model.
Final model	**Model C**: *response time ~ trait math anxiety + paradigm + (1 + paradigm | subject) + (1 | item)*	Final model = Step 3
Random effects	item (intercept, variance = 0.21, SD = 0.46) subject (intercept, variance = 0.28, SD = 0.53) paradigm (slope, variance = 0.13, SD = 0.37, r = -0.07)	

3. State (math) anxiety across three-time points

(1) LMM for state math anxiety across three-time points:

Full model	*state math anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + time^2 : trait math anxiety + (1 | subject)*	
Step 1	*state math anxiety ~ time + time^2 + trait math anxiety + time^2 : trait math anxiety + (1 | subject)*	The model excluding the interaction of time and trait math anxiety was significantly different from the original model (χ^2 = 24.40, p < .001). Therefore, we retained the interaction of time and trait math anxiety in the final model.
Step 2	*state math anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + (1 | subject)*	The model excluding the interaction of $time^2$ and trait math anxiety was not significantly different from the original model (χ^2 = 1.58, p = .209). Therefore, we removed the interaction of $time^2$ and trait math anxiety in the final model.
Step 3	*state math anxiety ~ time + trait math anxiety + time : trait math anxiety + (1 | subject)*	The model excluding the fixed effect of $time^2$ was not significantly different from the original model (χ^2 = 3.90, p = .142). Therefore, we removed the fixed effect of $time^2$ in the final model.

Final Model	**Model D**: *state math anxiety ~ time + trait math anxiety + time : trait math anxiety + (1 | subject)*	Final model = Step 3
Random effects	subject (intercept, variance = 0.26, SD = 0.51) (residual, variance = 0.22, SD = 0.47)	

(2) LMM for state anxiety across three-time points:

Full model	*state anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + time^2 : trait math anxiety + (1 | subject)*	
Step 1	*state anxiety ~ time + time^2 + trait math anxiety + time^2 : trait math anxiety + (1 | subject)*	The model excluding the interaction of time and trait math anxiety was significantly different from the original model ($\chi^2 = 14.52$, $p < .001$). Therefore, we retained the interaction of time and trait math anxiety in the final model.
Step 2	*state anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + (1 | subject)*	The model excluding the interaction of time² and trait math anxiety was not significantly different from the original model ($\chi^2 = 0.90$, $p = .344$). Therefore, we removed the interaction of time² and trait math anxiety in the final model.
Step 3	*state anxiety ~ time + trait math anxiety + time : trait math anxiety + (1 | subject)*	The model excluding the fixed effect of time² was not significantly different from the original model ($\chi^2 = 1.92$, $p = .382$). Therefore, we removed the fixed effect of time² in the final model.
Final Model	**Model S1**: *state anxiety ~ time + trait math anxiety + time : trait math anxiety + (1 | subject)*	Final model = Step 3
Random effects	subject (intercept, variance = 0.16, SD = 0.40) (residual, variance = 0.17, SD = 0.41)	

4. State (math) anxiety across six-time points

(1) LMM for state math anxiety across six-time points:

Full model	*state math anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + time^2 : trait math anxiety+ (1 | subject) + (1 | paradigm)*	
Step 1	*state math anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + time^2 : trait math anxiety+ (1 | subject)*	The model excluding the random intercept of paradigm was significantly different from the original model ($\chi^2 = 80.18$, $p < .001$). Therefore, we retained the random intercept of paradigm in the final model.

Step 2	*state math anxiety ~ time + time^2 + trait math anxiety + time^2 : trait math anxiety+ (1 | subject) + (1 | paradigm)*	The model excluding the interaction of time and trait math anxiety was significantly different from the original model (χ^2 = 15.45, p < .001). Therefore, we retained the interaction of time and trait math anxiety in the final model.
Step 3	*state math anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + (1 | subject) + (1 | paradigm)*	The model excluding the interaction of time^2 and trait math anxiety was not significantly different from the original model (χ^2 = 0.63, p = .428). Therefore, we removed the interaction of time^2 and trait math anxiety in the final model.
Step 4	*state math anxiety ~ time + trait math anxiety + time : trait math anxiety + (1 | subject) + (1 | paradigm)*	Remove the main effect of time^2. ANOVA showed it was not significant (χ^2 = 0.81, p = .668). Removed. The model excluding the main effect of time^2 was not significantly different from the original model (χ^2 = 0.81, p = .668). Therefore, we removed the main effect of time^2 in the final model.
Final model	**Model E**: *state math anxiety ~ time + trait math anxiety + time : trait math anxiety + (1 | subject) + (1 | paradigm)*	Final model = Step 4
Random effects	subject (intercept, variance = 0.33, SD = 0.57) paradigm (intercept, variance = 0.06, SD = 0.24) (residual, variance = 0.17, SD = 0.41)	

(2) LMM for state anxiety across six-time points:

Full model	*state anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + time^2 : trait math anxiety + (1 | subject) + (1 | paradigm)*	
Step 1	*state anxiety ~ time + time^2 + trait math anxiety + time : trait math anxiety + time^2 : trait math anxiety + (1 | subject)*	The model excluding the random intercept of paradigm was significantly different from the original model (χ^2 = 108.58, p < .001). Therefore, we retained the random intercept of paradigm in the final model.
Step 2	*state anxiety ~ time + time^2 + trait math anxiety + time^2 : trait math anxiety + (1 | subject) + (1 | paradigm)*	The model excluding the interaction of time and trait math anxiety was not significantly different from the original model (χ^2 = 1.90, p = .168). Therefore, we removed the interaction of time and trait math anxiety in the final model.
Step 3	*state anxiety ~ time + time^2 + trait math anxiety +*	The model excluding the interaction of

	(1 | subject) + (1 | paradigm)	$time^2$ and trait math anxiety was significantly different from the original model ($\chi^2 = 6.33$, $p = .042$). Therefore, we retained the interaction of $time^2$ and trait math anxiety in the final model.
Step 4	*state anxiety ~ time^2 + trait math anxiety + time^2 : trait math anxiety + (1 | subject) + (1 | paradigm)*	Remove the main effect of time. The model excluding the main effect of time was significantly different from the original model ($\chi^2 = 55.30$, $p < .001$). Therefore, we retained the main effect of time in the final model.
Final model	**Model S2:** *state anxiety ~ time + time^2 + trait math anxiety + time^2 : trait math anxiety + (1 | subject) + (1 | paradigm)*	Final model = Step 2
Random effects	subject (intercept, variance = 0.27, SD = 0.52) paradigm (intercept, variance = 0.08, SD = 0.28) (residual, variance = 0.16, SD = 0.40)	

5. Anxiety-complexity effect for response time but not accuracy

(1) LMM for response time:

Full model	*response time ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1 + paradigm | subject) + (1 | item)*	
Step 1	*response time ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1| subject) + (1 | item)*	The model excluding the random slope for paradigm was significantly different from the original model ($\chi^2 = 722.34$, $p < .001$). Therefore, we retained the random slope for paradigm in the final model.
Step 2	*response time ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1 + paradigm | subject)*	The model excluding the random intercept for item was significantly different from the original model ($\chi^2 = 1748.60$, $p < .001$). Therefore, we retained the random intercept for item in the final model.
Step 3	*response time ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1 + paradigm | subject) + (1 | item)*	The model excluding the interaction of trait math anxiety and difficulty was significantly different from the original model ($\chi^2 = 29.77$, $p < .001$). Therefore, we retained the interaction of trait math anxiety and difficulty in the final model.
Final Model	**Model S3**: *response time ~ trait math anxiety +*	Final model = Full model

	difficulty + trait math anxiety : difficulty + (1 + paradigm | subject) + (1 | item)	
Random effects	item (intercept, variance = 0.28, *SD* = 0.53) subject (intercept, variance = 0.65, *SD* = 0.81) paradigm (slope, variance = 0.33, *SD* = 0.57, *r* = .06) (residual, variance = 2.10, *SD* = 1.45)	

(2) GLMM for accuracy:

Full model	*accuracy ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1 + paradigm | subject) + (1 | item)*	
Step 1	*accuracy ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1 | subject) + (1 | item)*	The model excluding the random slope for paradigm was significantly different from the original model (χ^2 = 78.40, p < .001). Therefore, we retained the random slope for paradigm in the final model.
Step 2	*accuracy ~ trait math anxiety + difficulty + trait math anxiety : difficulty + (1 + paradigm | subject)*	The model excluding the random intercept for item was significantly different from the original model (χ^2 = 97.85, p < .001). Therefore, we retained the random intercept for item in the final model.
Step 3	*accuracy ~ trait math anxiety + difficulty + (1 + paradigm | subject) + (1 | item)*	The model excluding the interaction of trait math anxiety and difficulty was not significantly different from the original model (χ^2 = 0.36, p = .550). Therefore, we removed the interaction of trait math anxiety and difficulty in the final model.
Step 4	*accuracy ~ difficulty + (1 + paradigm | subject) + (1 | item)*	The model excluding the main effect of trait math anxiety was not significantly different from the original model (χ^2 = 4.48, p = .107). Therefore, we removed the main effect of trait math anxiety in the final model.
Step 5	*accuracy ~ (1 + paradigm | subject) + (1 | item)*	The model excluding the main effect of difficulty was significantly different from the original model (χ^2 = 93.59, p < .001). Therefore, we retained the main effect of difficulty in the final model.
Final Model	**Model S4**: *accuracy ~ difficulty + (1 + paradigm | subject) + (1 | item)*	Final model = Step 4
Random	item (intercept, variance = 0.16, SD = 0.40)	

effects	subject (intercept, variance = 0.30, SD = 0.55) paradigm (slope, variance = 0.22, SD = 0.46, r = -0.11)	

6. No speed-accuracy trade-off

Full model	*accuracy ~ paradigm + trait math anxiety + response time + paradigm : trait math anxiety + paradigm: response time + trait math anxiety : response time + paradigm : trait math anxiety : response time + (1 | subject)*	
Step 1	*accuracy ~ paradigm + trait math anxiety + response time + paradigm : trait math anxiety + paradigm: response time + trait math anxiety : response time + (1 | subject)*	The model excluding the three-way interaction was not significantly different from the original model (χ^2 = 0.30, p = .585). Therefore, we removed the three-way interaction in the final model.
Step 2	*accuracy ~ paradigm + trait math anxiety + response time + paradigm : trait math anxiety + trait math anxiety : response time + (1 | subject)*	The model excluding the interaction between paradigm and response time was not significantly different from the original model (χ^2 = 1.66, p = .435). Therefore, we removed the interaction between paradigm and response time in the final model.
Step 3	*accuracy ~ trait math anxiety + response time + trait math anxiety : response time + (1 | subject)*	The model excluding the interaction between paradigm and trait math anxiety was not significantly different from the original model (χ^2 = 1.75, p = .625). Therefore, we removed the interaction between paradigm and trait math anxiety in the final model.
Step 4	*accuracy ~ paradigm + trait math anxiety + response time + (1 | subject)*	The model excluding the interaction between trait math anxiety and response time was not significantly different from the original model (χ^2 = 4.38, p = .357). Therefore, we removed the interaction between trait math anxiety and response time in the final model.
Step 5	*accuracy ~ trait math anxiety + response time + (1 | subject)*	The model excluding the main effect of paradigm was significantly different from the original model (χ^2 = 25.22, p < .001). Therefore, we retained the main effect of paradigm in the final model.
Step 6	*accuracy ~ paradigm + response time + (1 | subject)*	The model excluding the main effect of trait math anxiety was not significantly

		different from the original model (χ^2 = 6.49, p = .262). Therefore, we removed the main effect of trait math anxiety in the final model.
Step 7	*accuracy ~ paradigm + (1 | subject)*	The model excluding the main effect of response time was not significantly different from the original model (χ^2 = 7.91, p = .245). Therefore, we removed the main effect of response time in the final model.
Final model	**Model S5**: *accuracy ~ paradigm + (1 | subject)*	Final model = Step 7
Random effects	subject (intercept, variance = 0.15, SD = 0.39) (residual, variance = 0.08, SD = 1.45)	

Part 3: Exploratory analysis

Exploratory analysis 1: Correlation between state math anxiety and performance across paradigms

To examine whether performance might account for the observed paradigm effects on state math anxiety (SMA), we computed *Pearson* correlations between SMA and arithmetic performance, including response time (RT) and accuracy (ACC), across and within paradigms. As shown in Table S3, changes in SMA (ΔSMA) from decision to production paradigms were not significantly correlated with changes in RT, but showed a significant negative correlation with changes in ACC. This indicates that individuals who show higher increases of state math anxiety in production than decision paradigms are the ones who make more errors in production than decision paradigms. Within paradigms, SMA correlated negatively with accuracy but not with RT in production paradigms, whereas no significant correlations were found for decision paradigms. This indicates that higher state math anxiety is related to more errors in production paradigms. This pattern suggests that heightened SMA is more closely related to concerns about errors rather than slower processing.

Table S4. Exploratory correlation analyses between state math anxiety and performance across and within paradigms

Correlations	***r***	***95% CI***	***p***
ΔSMA and ΔRT	.18	[-.07, .41]	.153
ΔSMA and ΔACC	-.27	[-.49, -.03]	**.028**
SMA and RT (production)	.23	[-.02, .45]	.070
SMA and ACC (production)	-.26	[-.48, -.02]	**.034**
SMA and RT (decision)	.17	[-.08, .40]	.179
SMA and ACC (decision)	-.08	[-.32, .16]	.506

Exploratory analysis 2: Time-wise correlation between state math anxiety and performance

To further explore how the relationship between state math anxiety (SMA) and task performance evolved over time, we computed time-wise *Pearson* correlations between SMA and response time (RT) as well as logit-transformed accuracy (ACC) across the six measurement time points. For each time point, the correlation coefficient (r) and its 95% confidence interval were estimated using Fisher's z transformation (see Table S4 and Figure S3).

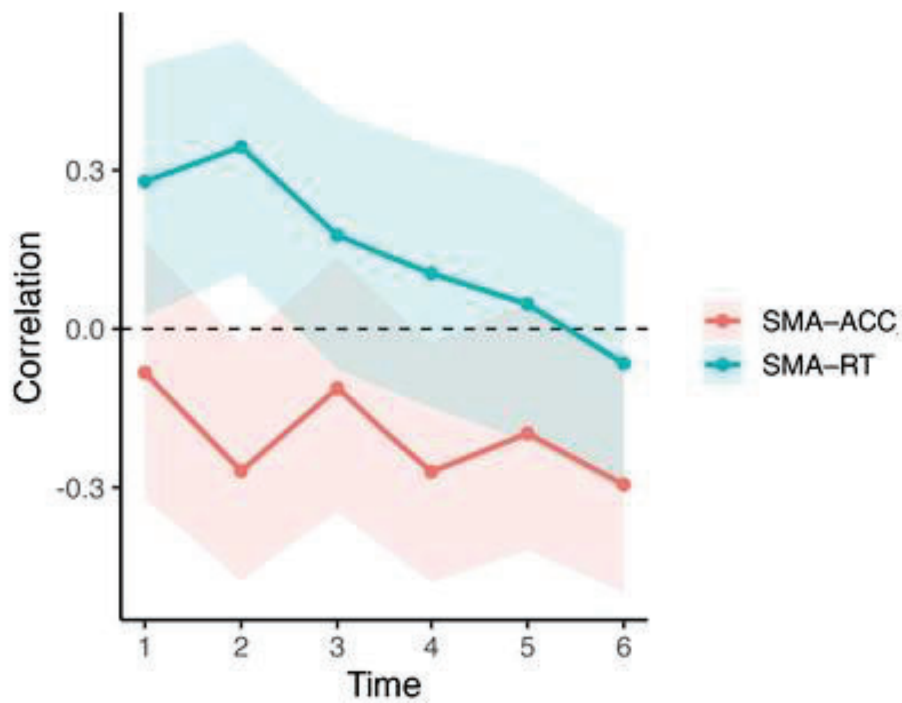
Overall, higher SMA tended to be associated with slower responses (positive SMA and RT correlation) and lower accuracy (negative SMA–ACC correlation). For RT, the strength of these associations decreased over time, suggesting that the coupling between anxiety and performance was strongest at the beginning of the experiment and attenuated as participants became more familiar with the task.

These exploratory findings indicate that state math anxiety exerts a transient impact on early performance before participants adapt to task demands.

Table S5. Correlation between state math anxiety and performance across six time points

Correlations	**Time**	***r***	***95% CI***	***p***
SMA and RT	1	.28	[.024, .499]	**.033**
	2	.34	[.105, .545]	**.006**
	3	.18	[-.076, .409]	.169
	4	.11	[-.149, .346]	.417
	5	.05	[-.212, .299]	.725
	6	-.07	[-.308, .185]	.610
SMA and ACC	1	-.08	[-.322, .166]	.516
	2	-.27	[-.481, -.027]	**.030**
	3	-.11	[-.348, .137]	.378
	4	-.27	[-.484, -.026]	**.031**
	5	-.20	[-.423, .050]	.117
	6	-.29	[-.502, -.054]	**.017**

Figure S3. Time-wise correlation between state math anxiety and performance



Notes. ACC represents accuracy, RT represents response time.

A3. Study 3

A3.1 Publication of Study 3

Yao, X., Avcil, M., Meuer, P., Nuerk, H.-C., & Artemenko, C. Math self-concept decreases while math anxiety increases over the lifespan. Annals of the New York Academy of Sciences. (in revision)

Math self-concept decreases while math anxiety increases over the lifespan

Xinru Yao[1, 2,*], Mine Avcil[1], Paul Meuer[1], Hans-Christoph Nuerk[1, 2, 3], Christina Artemenko[1, 2, 3,*]

[1] Department of Psychology, University of Tuebingen, Tuebingen, Germany

[2] LEAD Graduate School and Research Network, University of Tuebingen, Tuebingen, Germany

[3] German Center for Mental Health (DZPG), Tuebingen, Germany

* Correspondence concerning this article should be addressed to Xinru Yao or Christina Artemenko:

Name: Xinru Yao

Email address: xinru.yao.tuebingen@gmail.com

Name: Christina Artemenko

Email address: christina.artemenko@uni-tuebingen.de

Author contributions: CRediT

Conceptualization: Xinru Yao, Hans-Christoph Nuerk, Christina Artemenko

Methodology: Xinru Yao, Mine Avcil, Hans-Christoph Nuerk, Christina Artemenko

Software: Xinru Yao, Paul Meuer

Validation: Xinru Yao, Christina Artemenko

Formal analysis: Xinru Yao

Investigation: Mine Avcil

Resources: -

Data Curation: Xinru Yao

Writing – original draft: Xinru Yao, Paul Meuer

Writing – review & editing: Xinru Yao, Mine Avcil, Paul Meuer, Hans-Christoph Nuerk, Christina Artemenko

Visualization: Xinru Yao, Paul Meuer

Supervision: Christina Artemenko, Hans-Christoph Nuerk

Project Administration: Xinru Yao

Funding Acquisition: Christina Artemenko

Abstract

Math self-concept and math anxiety are two critical psychological constructs influencing mathematical performance. Math self-concept reflects an individual's perception of their mathematical abilities, while math anxiety represents a negative emotional response to mathematical tasks. Although extensively studied in young adults and children, little is known about how these constructs evolve across the lifespan. Findings from young adults are sometimes generalized to all adults, suggesting an assumption of stability over time. However, given that other personality traits exhibit significant variation across the lifespan, this assumption may not hold.

This study systematically examined the developmental trajectories of math self-concept and math anxiety, their interrelationship, and their impact on arithmetic performance from childhood to older adulthood. The findings revealed a decline in math self-concept and a concurrent increase in math anxiety with age. Additionally, a strong negative correlation between math self-concept and math anxiety persisted across all age groups. Importantly, math self-concept, rather than math anxiety, predicted arithmetic performance as well as the complexity effect.

These results provide critical insights into the lifelong development of math-related psychological constructs, suggesting that findings from young adults cannot be generalized across developmental stages. The observed increase in math anxiety and decline in math self-concept with age indicate progressively more problematic attitudes and emotions towards mathematics.

Keywords

math self-concept, math anxiety, arithmetic, lifespan development, children, aging

Introduction

Math performance is related to both cognitive factors and personality traits, particularly to math self-concept and math anxiety. Math self-concept, i.e., the perception of one's own mathematical ability, is strongly linked to academic success, predicting school grades and long-term engagement with mathematics (Marsh et al., 2005; Möller et al., 2020). Math anxiety, on the other hand, can undermine performance and leads to avoidance of math-related activities and careers, particularly in STEM fields (Chipman et al., 1992; Meece et al., 1990). The PISA data showed a high correlation between math self-concept and math performance (Ding et al., 2024). The 2022 PISA study reported that nearly half of students felt anxious about struggling in math class or receiving poor grades, with 15% experiencing anxiety during lessons (OECD, 2023). This concerning educational trend raises the question of how these patterns evolve across the lifespan. Is math-related anxiety merely a transient phenomenon that fades in adulthood, or does it persist and even intensify once formal mathematical education ends?

Despite extensive research on these constructs, most studies focused on specific age groups, such as school-aged children or university students. It is often implicitly assumed that what we observe in university students allows for general statements about the relationship of math attitude, anxiety and performance across adulthood. This assumption may be questionable, because we know that other personality traits change over the lifespan (Asendorpf, 1999), especially when people grow older, their stability decreases (Wortman et al., 2012). Therefore, the current study sets out to systematically examine the development of math self-concept, math anxiety, and their relation to math performance in 6 different age groups, providing a comprehensive perspective on their interplay over the lifespan.

Lifespan development of math self-concept

Math self-concept refers to an individual's perception of their mathematical abilities and competencies, both in academic and everyday contexts (Erdogan & Sengul, 2014; Möller & Trautwein, 2009). It is not an objective evaluation of mathematical ability but is shaped by external influences such as cultural expectations and social comparisons (Ertl et al., 2017). While research on the development of math self-concept has primarily focused on children and adolescents, its development in older adults was not studied yet.

When children start school, they generally exhibit a high math self-concept (Lindberg et al., 2013). This is largely due to their limited academic experience and minimal social comparisons, leading to an optimistic view of their abilities. During this stage, evaluations are often learning-focused rather than performance-based, which helps to sustain a positive self-concept (Fredricks & Eccles, 2002).

As younger adolescents progress through primary and early secondary school, their math self-concept begins to decline. This decline is supported by both longitudinal and cross-sectional studies (Fredricks & Eccles, 2002; Marsh, 1989; Nagy et al., 2010). The transition to secondary school introduces increased exposure to academic comparisons and more competitive evaluation systems, further accelerating this decline (Lehikoinen et al., 2024).

For older adolescents during late secondary or high school, the decline in math self-concept becomes more evident. Social comparisons play a significant role, as students become increasingly aware of their academic standing relative to peers (Helmke, 1999; Nagy et al., 2010). This period also coincides with critical educational transitions and heightened academic pressure, contributing to a sharper drop in math self-concept, particularly between grades 9 and 10 (ages 14-16 years, Nagy et al., 2010).

Upon entering adulthood, math self-concept seems to stabilize. Marsh (1989) found no significant differences in math self-concept between individuals aged 18–20.5 and those over 21.5 years old. Similarly, studies on adult learners aged 18–59 years reported no age-related changes in math self-concept (Jameson & Fusco, 2014; Skaalvik & Skaalvik, 2005), suggesting that individuals develop a consistent perception of their mathematical abilities by early adulthood. However, these findings are largely based on samples dominated by young adults (e.g., mostly university students), with minimal consideration of older adults.

Despite its relevance in financial decision-making and daily numeracy, there is a lack of research on math self-concept in older adulthood. Given the age-related changes in many cognitive domains as well as the increasing interindividual differences during aging (Glisky, 2007; Hultsch et al., 2008), it is crucial to understand how math self-concept changes in later life. The lack of lifespan studies covering all age groups highlights a major research gap in this topic that needs to be addressed.

In summary, math self-concept declines throughout childhood and adolescence but may stabilize in adulthood. Although existing research provides valuable insights, comprehensive studies spanning the entire lifespan, particularly including older adulthood, are still needed to fully understand the developmental trajectory of math self-concept.

Lifespan development of math anxiety

Math anxiety refers to the negative emotional response individuals experience when engaging with numerical or mathematical tasks (Ashcraft, 2002; Richardson & Suinn, 1972). It manifests through tension, worry, nervousness, and fear in both academic and everyday situations (Hembree, 1990; Hopko et al., 2002).

Math anxiety can emerge early in childhood. While children generally report low to moderate levels of math anxiety (Szczygieł, 2019), it has been detected even in kindergarteners (Lu et al., 2021). Studies show that math anxiety increases in primary school between grades 1 and 3 (Krinzinger et al., 2009), despite children having limited exposure to math-related performance situations. This suggests that early experiences and environmental factors play a key role in the development of math anxiety.

Math anxiety intensifies significantly during adolescence, with research reporting a sharp increase between grades 6 and 8 (Birgin et al., 2010). According to Hembree's (1990) meta-analysis, math anxiety peaks in grades 9 and 10 (ages 14–16 years). This period coincides with increased academic pressure, social comparisons, and performance evaluations, which are likely to intensify anxiety. However, Sorvo et al. (2019) suggest that math anxiety may decline slightly in older children, particularly between grades 2 and 5, pointing at variability across age groups and studies.

For adults, math anxiety tends to decrease but remains prevalent. Hembree (1990) found that math anxiety stays stable in younger adults after school. Hart and Ganley (2019) found that math anxiety persists in middle-aged and older adults, with only a slight downward trend observed across adulthood ($r = -.06$, $p = .082$). Supporting this age-related trend, older adults were found to report lower levels of math anxiety compared to younger adults While there is first evidence that math anxiety persists during adulthood, further evidence is needed to test the assumption that math anxiety declines during adulthood and thus only might be a transient phenomenon during education.

In summary, math anxiety is present in all age groups across the lifespan, pointing at the relevance for further investigation. Math anxiety generally increases throughout childhood and adolescence, peaks in secondary school, and then might slightly decline in adulthood. Critically, research on aging is limited, making it challenging to determine how math anxiety develops across the lifespan.

Reciprocal relationship between math self-concept and math anxiety

The developmental changes in math self-concept and math anxiety seem to follow similar trajectories. This is because they are distinct but closely related constructs. While math self-concept refers to cognitive components, reflecting one's perception of mathematical abilities, math anxiety refers to affective components, reflecting emotional reactions to math-related tasks (Klee et al., 2022). Studies have consistently demonstrated a strong negative correlation between math self-concept and math anxiety in adolescents and younger adults, with a correlation of $r = -.71$ according to a meta-analysis (Hembree, 1990). However, the samples may not be representative for all age groups.

Individuals with low math self-concept tend to doubt their mathematical abilities, feel inadequate in math-related situations, and consequently experience higher levels of math anxiety. Conversely, math anxiety can undermine math self-concept: Anxiety distorts self-perception, causing individuals with high math anxiety to undervalue their math abilities and avoid math-related tasks, which further reinforces negative self-beliefs. Hence, the relationship between math self-concept and math anxiety is reciprocal since math self-concept predicted math anxiety and vice versa over a one-year period (Ahmed et al., 2012; Pekrun, 2006). The cumulative stability concept of personality development (Roberts & Caspi, 2003) suggests that personality traits and their environment mutually reinforce one another over time. If math self-concept and math anxiety stabilize in this manner, then early educational experiences may have long-term consequences for math-related emotions and engagement later in life. To fully understand the lifelong dynamics of math self-concept and math anxiety, research need to address this interplay across the entire lifespan – not only in childhood and adolescence but also during adulthood.

Arithmetic performance is related to math self-concept and math anxiety

Math self-concept plays a critical role in influencing math performance (Arens et al., 2017; Marsh et al., 2018; Perinelli et al., 2022). A meta-analysis by Möller et al. (2009) reported a moderate positive correlation ($r = .43$) between math self-concept and math performance.

The Skill-Development Model suggests that academic performance shapes self-concept, as students assess their abilities by comparing their past performance and that of their peers based on comparisons with their own past performance and with others (Marsh, 1990). For example, students in high-achieving environments often develop lower academic self-concepts than their peers in low-achieving settings, a phenomenon known as the Big-Fish-Little-Pond Effect (Craven & Marsh, 2000; Seaton et al., 2010). Conversely, the Self-Enhancement Approach posits that self-concept predicts future academic performance by fostering motivation, effort, interest, and self-efficacy (Marsh & Yeung, 1997). Convergently, the Reciprocal-Effects Model argues that academic self-concept and performance mutually influence and reinforce one another over time. Empirical evidence by longitudinal studies and meta-analyses support this model, showing that prior self-concept predicts later academic achievement, while past achievement also influences subsequent self-concept (Arens et al., 2017; Marsh et al., 2018; Valentine et al., 2004).

This relation between math self-concept and math performance further strengthens with age in early development (Guay et al., 2003; Möller et al., 2020). Guay et al. (2003) observed that for students at the end of primary school, academic self-concept becomes increasingly stable over three years and shows a stronger correlation with academic performance. Similarly, a longitudinal study found that the relation between math self-concept and performance becomes stronger from the age of 10 to 13 years (Perinelli et al., 2022). The meta-analysis by Möller et al. (2020) further supports these findings, highlighting that the relation between math self-concept and performance intensifies over time, particularly during the transition to secondary school. Afterwards, this relation remains consistent and stable in young and middle adulthood (Skaalvik & Skaalvik, 2005). However again, lifetime development up to older adulthood was never investigated.

Math anxiety negatively affects math performance in various areas, including basic numerical understanding, arithmetic, word problem solving, fractions, geometry, algebra, and statistics (Barroso et al., 2021). Ramirez et al. (2018) summarized three theories explaining the relation between math anxiety and performance: The Disruption Account argues that math anxiety triggers intrusive thoughts and ruminations, which occupy working memory resources, reduce cognitive efficiency and subsequently impair math performance. The Reduced Competency Account suggests that math anxiety is the result of poor math ability, where reduced competency leads to difficulties in learning and performance, ultimately causing anxiety. The Interpretation Account holds that math anxiety stems not only from poor math skills or negative experiences but from how individuals interpret and appraise their math-related experiences. Finally, – similar to math self-concept – researchers also suggest a bidirectional relationship between math anxiety and math performance, resulting in a vicious cycle (Carey et al., 2016; Cipora et al., 2022).

The negative relationship between math anxiety and math performance begins and strengthens during school years (Barroso et al., 2021; Namkung et al., 2019). Some studies observed this negative relation already in young children during the early school years (Harari et al., 2013; Ramirez et al., 2013; Wu et al., 2012), while others only found it in adolescents but not in children (Hill et al., 2016; Krinzinger et al., 2009). Meta-analyses further indicate that the correlation becomes stronger with age, increasing from childhood ($r = -.27$) to adolescence ($r = -.36$) (Namkung et al., 2019) and from grade 3 ($r = -.20$) to grade 12 ($r = -.34$) (Barroso et al., 2021).

Although this negative relationship has been reported in adulthood, most existing studies have focused on college students ($r = -.24$) or non-student adult samples ($r = -.32$) (Barroso et al., 2021). Limited research exists on older adults, with Skagerlund et al. (2018) being one of the few studies demonstrating a significant negative correlation ($r = -.32$) between math anxiety and numeracy in middle-aged adults ($\approx$ 50 years). However, meaningful comparisons across the lifespan remain difficult due to differences in sample characteristics, performance measures, and cultural contexts across studies. Thus, while existing findings suggest that the relationship between math anxiety and performance persists over the lifespan, there is a need for a lifespan study investigating different age groups with the same measures to better understand how the relation of math anxiety and performance change during development.

Finally, research has shown that math anxiety affects arithmetic performance particularly when arithmetic gets more complex (known as the anxiety-complexity effect; Ashcraft & Faust, 1994; Huber & Artemenko, 2021). This was found for arithmetic requiring the carry operation (e.g., 45 + 28, where the sum of units exceeds 9) or the borrow operation (e.g., 61 – 47, where the unit of the subtrahend is larger than the unit of the minuend), which impose greater cognitive demands in working memory, leading to worse performance (Artemenko et al., 2018; Imbo et al., 2007). Because their working memory capacity is constrained by intrusive anxious thoughts (Ashcraft & Faust, 1994; Ashcraft & Kirk, 2001), individuals with higher math anxiety perform particularly worse in arithmetic requiring carrying or borrowing (Huber & Artemenko, 2021; Yao, Huber, et al., 2025). Previous research has established the anxiety-complexity effect in math anxiety; however, its relevance to math self-concept and its developmental trajectory remain unexplored. Experimental research on emotion induction indicates that negative emotions might impact the performance in complex arithmetic in younger adults more than in older adults (Lallement & Lemaire, 2021; See also Lemaire, 2024).

Current Study

Building on the outlined developmental trajectories of math self-concept, math anxiety, and their influence on arithmetic performance, this preregistered study aims to systematically explore these constructs across the lifespan (see preregistration at https://aspredicted.org/g6hh4.pdf). Specifically, it investigates how math self-concept and math anxiety evolve with age, how they relate to arithmetic performance during the lifespan, and how these relationships are shaped by the complexity of arithmetic:

H1: *Age-dependent changes of concepts*. (H1a) Math self-concept is expected to decrease from childhood to adulthood and remain stable during adulthood. (H1b) Math anxiety is expected to increase from childhood to adolescents and to decrease during adulthood.

H2: *Relations between concepts*. (H2a) A positive relation between math self-concept and arithmetic performance is expected within and across age groups. (H2b) A negative relation between math anxiety and arithmetic performance is expected within and across age groups. (H2c) A negative relation between math self-concept and math anxiety is expected within and across age groups.

H3: *Age-dependent changes of relations between concepts*. We expect age-dependent changes in the relations of math self-concept and math anxiety with arithmetic performance. (H3a) The relation between math self-concept and arithmetic performance is expected to increase from primary school age over secondary school age to high school age and remain stable during adulthood. (H3b) The relation between math anxiety and arithmetic performance is expected to increase from primary school age over secondary school age to high school age and decrease during adulthood.

H4: *Interactions with arithmetic complexity*. Interactions of arithmetic complexity with math self-concept and math anxiety are expected within and across age groups, i.e., larger carry/borrow effects are expected to be associated with (H4a) lower math self-concept and (H4b) higher math anxiety. (H4c) We expect these interactions to change across the lifespan.

Methods

Participants

In a cross-sectional study, 306 participants were recruited with a broad age range from children in the third grade of primary school to older adults (see Table 1; sample taken from Avcil & Artemenko, 2023). The sample was categorized into six age groups: children (8–10 years old, grades 3–4), younger adolescents (10–14 years old, grades 5–8), older adolescents (14–18 years old, grades 9–13), younger adults (18–34 years old), middle-aged adults (35–59 years old), and older adults (60 years and older). All participants were either native German speakers or enrolled in the German school system. The study was approved by the Ethics Committee for Psychological Research of the University of Tuebingen.

Materials

Math self-concept. Math self-concept was assessed using the adapted German version of the mathematics subscale (Schwanzer et al., 2005) from the Self-Description Questionnaire III (SDQ III; Marsh, 1990). This subscale includes four items, such as "I am good at math." and participants rate their agreement with these statements on a 4-point Likert scale (from 1 = strongly disagree to 4 = strongly agree). The mean score of the four items was calculated, with Items 2 and 4 being reverse-coded, with higher scores indicating a stronger math self-concept (*Theoretical Range*: 1–4). The German version of the SDQ III math subscale has demonstrated very good reliability (Cronbach's Alpha = .89; Schwanzer et al., 2005). In our study, the overall Cronbach's Alpha was .88 (.72 in children; .82 in younger adolescents; .88 in older adolescents; .91 in younger adults; .90 in middle-aged-adults; .91 in older adults).

Math anxiety. Math anxiety was measured using the German version of the Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003), which previously showed very good internal reliability (Cronbach's $\alpha = 0.89$) and test-retest reliability ($r = .85$) (Artemenko et al., 2021). The questionnaire consists of 9 items describing math-related situations with a 5-point Likert scale (from 1 = low anxiety to 5 = high anxiety). The mean score of the items was calculated with higher scores indicating higher math anxiety (*Theoretical Range*: 1–9). In our study, the overall Cronbach's Alpha was .87 (.64 in children; .71 in younger adolescents; .77 in older adolescents; .87 in younger adults; .92 in middle-aged-adults; .93 in older adults).

Arithmetic task. The arithmetic task included 128 addition and subtraction problems with and without carry or borrow operation in equal parts (for details see Avcil & Artemenko, 2023). All arithmetic problems included two-digit operands resulting in a two-digit solution. The task was presented using OpenSesame (Mathôt et al., 2012). In a computerized written production paradigm (Yao et al., 2025), participants were instructed to solve the problems as quickly and accurately as possible and type the response into a number keyboard. Each trial began with a fixation on the screen for 500 ms, then the problems appeared without a time limit until the participant responded. Each trial ended with a black screen for 500 ms after the participants' response, then the next trial came. Moreover, we used averages across all conditions for reaction time (RT) and logit-transformed error rate (ER; the Shapiro-Wilk test confirmed that the logit-transformed error rate data followed a normal distribution) as dependent variables for arithmetic performance.

Procedures

Participants first completed the arithmetic task and a number comparison task (the order of the two tasks was counterbalanced and the other task is not relevant for the current study). Next, they performed cognitive tests designed to assess inhibition, working memory, and processing speed, which will not be analyzed in the current study (cf. Avcil & Artemenko, 2023). Finally, the math self-concept and math anxiety questionnaires were administered.

Analysis

Data exclusion. As preregistered, participants were excluded from all analyses, if they do not meet the inclusion criteria (age, German language, no cognitive impairment). To ensure cognitive functioning in older adults, all participants over 60 years of age completed the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). *Z*-scores were calculated using normative data that accounted for age, gender, and education level (Thomann et al., 2018) and participants with a *z*-score below -1 (1 *SD* below the average) were excluded. Participants were case-wise excluded from the arithmetic task, if missing data exceeds 50% of the task. Additionally, participants were also case-wise excluded from each task, if they deviated more than 3 median absolute deviations from the groups' median reaction time (RT) of the arithmetic task, and if less than 25% of valid RT data per condition remained for RT analysis of the arithmetic task. Finally, participants who did not fully complete the math anxiety and math self-concept questionnaires were excluded. On trial level, the analysis of reaction time (RT) only included RTs (1) from correctly solved trials (2) longer than 0.2 s, and (3) that did not deviate more than 3 median absolute deviation from the individual median. After applying the exclusion criteria (for details see Supplementary Materials, Table S1), the final sample consisted of $N = 281$ (see Table 1).

Table 1. Descriptive statistics for each age group

	N	Age			Gender		Education
	incl (excl)	*M* (*SD*)	*Range*	m	f	d	*M* (*SD*)
Children	38 (15)	9.18 (0.69)	8–10	27	10	1	3.71 (0.46)
Younger adolescents	47 (3)	12.40 (1.14)	10–14	25	22	0	7.17 (0.94)
Older adolescents	50 (0)	15.42 (1.03)	14–17	18	32	0	9.84 (1.06)
Younger adults	53 (0)	23.83 (3.19)	19–33	12	40	1	16.58 (2.05)
Middle-aged adults	47 (3)	47.62 (8.14)	34–59	15	32	0	18.84 (3.70)
Older adults	46 (4)	71.91 (7.41)	60–88	21	27	0	16.13 (4.50)
Total	281 (25)	30.29 (22.77)	8–88	117	162	2	12.37 (5.93)

Notes. "incl" means the number of included participants and "excl" means the number of excluded participants. Age and education are given in years for the included participants. For gender, "m" means the number of male participants, "f" means the number of female participants and "d" means the number of diverse participants.

Statistical data analysis. The data was analyzed by Bayesian statistics using JASP (Jefferys's Amazing Statistics Program, Version 0,18, JASP Team, 2022). Bayesian statistics enable calculating graded evidence for the null hypothesis (H_0) and the alternative hypothesis (H_1; for detailed descriptions see Masson, 2011). Bayes factors (*BF*) indicate how much more likely the observed data will be under one compared to the other hypothesis (Faulkenberry et al., 2020).

The statistical analyses included Bayesian t-tests, Bayesian ANOVAs, and Bayesian regressions. The prior for *t*-tests was set to a Cauchy distribution centered on zero with an *r* width parameter of 0.707. The prior for ANOVAs was set to the default Cauchy prior of $r = 0.5$ for the fixed effects and to a uniform model prior. The prior for regressions was set to the default Jeffrey's–Zellner–Siow prior of $r = 0.354$ for regression coefficients and to a uniform model prior. For ANOVAs and regressions, each model was compared to the null model and Bayesian model averaging compared the models with the respective effect to equivalent models without the effect (as suggested by Sebastiaan Mathôt).

Beyond the preregistered analysis, we also included gender as a fixed factor to explore potential gender differences. The results revealed a gender difference for Hypothesis 1, while Hypotheses 3 and 4 produced similar outcomes regardless of gender (see Supplementary Materials 2 for details).

Results

All material, data, and analysis scripts are openly shared on OSF (https://osf.io/9tvxh/).

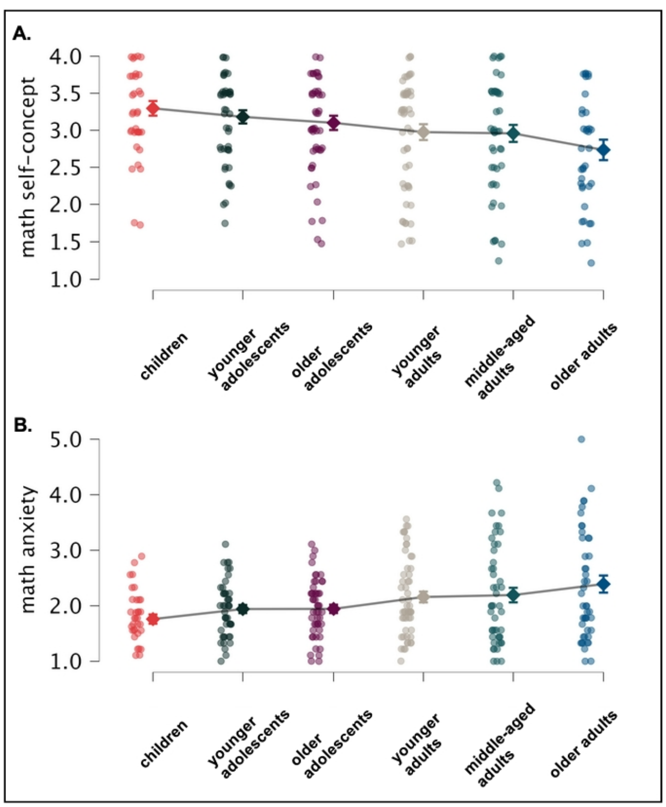
Age-dependent changes of math self-concept and math anxiety

For H1, Bayesian regressions on math self-concept and math anxiety were conducted with age group as a linear, logarithmic and quadratic predictor. Furthermore, Bayesian ANOVAs on math self-concept and math anxiety were conducted with age group as a between-subject factor.

For math self-concept, the regression favored the linear model (BF_{10} > 100), indicating strong evidence for a linear negative relation between age group and math self-concept (see Figure 1A; Supplementary Materials, Table S3). Math self-concept decreased over the lifespan, with a mean coefficient of -0.060, however, the evidence for this effect was inconclusive (BF_{incl} = 1.871). The ANOVA results for age group were inconclusive (BF_{10} = 1.77, BF_{01} = 0.57; see Supplementary Materials 1, Table S4).

For math anxiety, the regression favored the linear model (BF_{10} > 100), indicating strong evidence for a linear relation between age group and math anxiety (see Figure 1B; Supplementary Materials 1, Table S5). Math anxiety increased over the lifespan, with a mean coefficient of 0.081, however, the evidence for this effect was inconclusive (BF_{incl} = 2.35). The ANOVA results revealed strong evidence for the impact of age group on math anxiety (BF_{10} = 12.40). Post-hoc comparisons showed moderate or strong evidence that math anxiety was higher in all adult groups compared to children, and in older adults compared to both adolescent groups (see Supplementary Materials, Table S6).

Figure 1. Lifespan development of math self-concept and math anxiety



Notes. Linear trends for a decrease in math self-concept (A) and an increase in math anxiety (B) over the lifespan. Error bars represent the standard error of the mean (SE).

Relations between concepts

For H2, Bayesian correlations of math self-concept, math anxiety, arithmetic performance, and the arithmetic complexity effect were evaluated for each age group and across groups.

Math self-concept and math anxiety. Strong negative correlations between math self-concept and math anxiety were observed both within and across all age groups (see Table 2), indicating that a lower math self-concept was associated with higher levels of math anxiety, independent of age.

Math self-concept and arithmetic performance. The correlations between math self-concept and reaction time were inconclusive within all age groups, except for younger adults, where a lower math self-concept was associated with slower responses in arithmetic. Similarly, the correlations between math self-concept and error rate were inconclusive, except for older adolescents, where a lower math self-concept was associated with more errors in arithmetic.

Math anxiety and arithmetic performance. The correlations between math anxiety and arithmetic performance (reaction time and error rate) were inconclusive both within and across all age groups, except for older adults, where higher math anxiety was associated with slower responses in arithmetic.

Math self-concept and complexity effect. The correlations between math self-concept and complexity effect in terms of reaction time and error rate were inconclusive within all age groups, except for younger adults, where a lower math self-concept was associated with a larger complexity effect (reaction time & error rate) in arithmetic.

Math anxiety and complexity effect. The correlations between math anxiety and complexity effect (reaction time and error rate) were inconclusive both within and across all age groups, except for younger adults, where higher math anxiety was associated with a larger complexity effect (error rate) in arithmetic.

As there was no consistent pattern of correlations of math self-concept and math anxiety to performance, we further explored efficiency scores to address possible speed-accuracy trade-offs. However, the results were similar (see Supplementary Materials 3, Table S18).

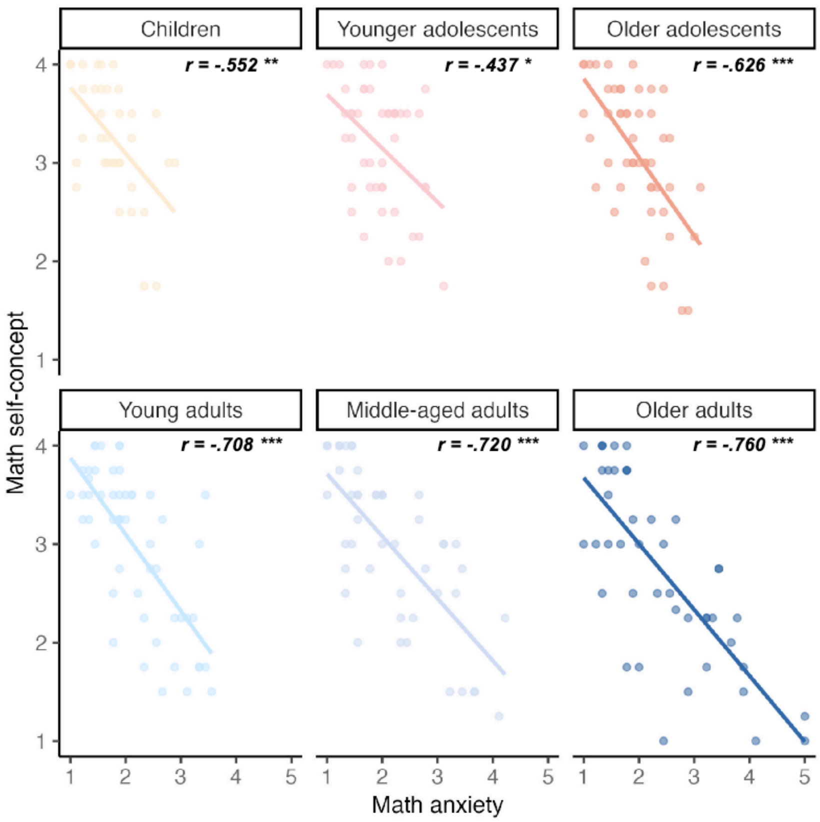
Table 2. Correlations within and across six age groups

	children	younger adolescents	older adolescents	younger adults	middle-aged adults	older adults	overall
math self-concept & math anxiety	-0.552 **	-0.437 *	-0.626 ***	-0.708 ***	-0.720 ***	-0.760 ***	-0.688 ***
math self-concept & reaction time	0.178	-0.123	-0.188	-0.445 **	-0.240	-0.239	-0.023
math self-concept & error rate	-0.259	-0.163	-0.491 **	-0.081	-0.120	-0.118	-0.083
math anxiety & reaction time	-0.130	0.169	-0.068	0.261	0.018	0.367 *	-0.026
math anxiety & error rate	0.249	-0.056	0.140	0.094	0.090	0.096	-0.013
math self-concept & complexity effect (RT)	0.199	-0.014	-0.015	-0.364 *	-0.218	-0.045	-0.083
math self-concept & complexity effect (ER)	0.039	-0.266	-0.023	-0.354 *	-0.344	-0.176	-0.179

math anxiety & complexity effect (RT)	-0.153	0.242	-0.233	0.316	0.062	0.111	-0.048
math anxiety& complexity effect (ER)	0.032	-0.007	-0.135	0.375 *	0.230	0.054	0.085

Notes. * $BF_{10} > 3$, ** $BF_{10} > 30$, *** $BF_{10} > 100$

Figure 2. Relations between math self-concept and math anxiety in each age group



Notes. Evidence supporting the negative correlation between math self-concept and math anxiety in each age group.

Age-dependent changes in relation of math self-concept and math anxiety to arithmetic performance

For H3, Bayesian ANCOVAs on arithmetic performance (error rate and reaction time) were conducted with age group as a between-subject factor, math self-concept and math anxiety as covariates, and their interactions with age group.[1]

[1] It was further planned to compare correlations pairwise between adjacent age groups; however, this analysis was not conducted due to the lack of correlations with arithmetic performance.

For error rate, the model including age group and math self-concept showed strong evidence (BF_{10} > 100; see Supplementary Materials 1, Table S7). Evidence was strong for the effect of age group (BF_{incl} > 100), indicating that middle-aged adults and children exhibited the highest error rates, whereas adolescents and younger adults performed more accurately (cf. Avcil & Artemenko, 2025). Evidence was moderate for the effect of math self-concept (BF_{incl} = 5.39), indicating that individuals with lower math self-concept showed higher error rates.

For reaction time, similarly, the model including age group and math self-concept showed the strongest evidence (BF_{10} > 100; see Supplementary Materials 1, Table S8). Evidence was strong for the effect of age group (BF_{incl} > 100), indicating that reaction time decreased from childhood to younger adulthood but increased again in middle-aged and older adults (cf. Avcil & Artemenko, 2025). However, evidence was inconclusive for the effect of math self-concept (BF_{incl} = 1.30, BF_{excl} = 0.75).

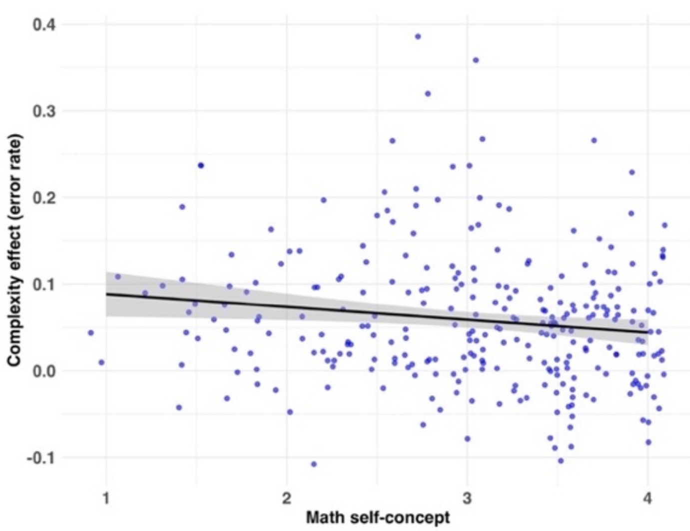
Interactions of arithmetic complexity with math self-concept and math anxiety

For H4, Bayesian ANCOVAs on the arithmetic complexity effect (error rate and reaction time) were conducted with age group as between-subject factor, math self-concept and math anxiety as covariates, and their interactions with age group.

For error rate, the model including age group and math self-concept showed strong evidence (BF_{10} = 71.91; see Supplementary Materials 1, Table S9). Evidence was moderate for the effect of age group (BF_{incl} = 3.74; cf. Avcil & Artemenko, 2025) and strong for the effect of math self-concept (BF_{incl} = 46.00), indicating that a lower math self-concept was associated with a larger arithmetic complexity effect (see Figure 2).

For reaction time, the model with age group only showed the strongest evidence (BF_{10} > 100; see Supplementary Materials 1, Table S11), and evidence was strong for the effect of age group (BF_{incl} > 100; cf. Avcil & Artemenko, 2025).

Figure 2. Math self-concept predicts arithmetic complexity effects



Discussion

The goal of this study was to investigate the development of math anxiety and math self-concept, as well as their relationship to math performance, across different age groups. We have three main findings: (1) Math self-concept decreases and math anxiety increases across lifespan. (2) Math self-concept and math anxiety are consistently negatively related across the lifespan. (3) Math self-concept can predict arithmetic performance (error rate) and the complexity effect (error rate).

Developmental trajectories of math self-concept and math anxiety

Interestingly, our data revealed a linear decline in math self-concept and a linear increase in math anxiety throughout the lifespan. It means that as individuals age, they become less confident in their math-related abilities while experiencing higher anxiety. This development confirms previous research on childhood and adolescence (e.g., Ahmed, 2018; Lehikoinen et al., 2024), but it contradicts our and others' expectations for adulthood. Our data strongly suggest that findings from younger adults can not simply be generalized to adulthood, as they may underestimate math anxiety and overestimate math self-concept in older adults. Which underlying factors can explain this pattern?

The internal/external frame of reference model suggests that the self-concept is shaped by both individual ability and comparative evaluations within different academic domains (Marsh, 1986). During development, children develop a clearer understanding of their strengths and weaknesses and their self-concept is increasingly influenced by their performance relative to peers. This may explain why math self-concept declines despite stable or improving mathematical skills (for arithmetic performance see Avcil & Artemenko, 2023). The decline in math self-concept may thus be driven by increasing social comparisons and performance evaluations, particularly during adolescence, when academic settings shift from learning oriented environments to more competitive and norm-referenced assessments (Fredricks & Eccles, 2002).

Similarly, the developmental increase in math anxiety from children to adults corroborates previous research (Birgin et al., 2010; Hembree, 1990; Krinzinger et al., 2009). As students progress through school, they face more complex math content, frequent testing, and increased academic competition. Moreover, adolescents and adults are more aware of the broader implications of math performance for academic and career prospects. Both of these amplify anxiety (Ashcraft & Krause, 2007; Dowker et al., 2016; Maloney & Beilock, 2012). On the other hand, negative experiences, such as repeated failure, discouraging feedback, or societal stereotypes, further reinforce math anxiety (Beilock et al., 2010; Ramirez et al., 2018).

Unexpectedly, the assumed stability of self-concept in adulthood obtained in some other studies (Jameson & Fusco, 2014; Skaalvik & Skaalvik, 2005) was not evident in our data. Instead, math self-concept continued to decline, while math anxiety increased—contrary to expectations that math-related attitudes would improve as adults move away from formal education and self-select into vocational paths that minimize their engagement with mathematics. One possible explanation for why prior studies found stability, while we observed change, lies in the age distribution of the samples used in previous research. Many studies on math self-concept and math anxiety in adults have primarily focused on younger and middle-aged adults, often excluding older adults (e.g., above 60 years) from their samples. However, research indicates that while personality traits and self-concept undergo notable changes in young adulthood (ages 20–40 years), they continue evolving into middle and old age (Roberts & Mroczek, 2008), so that restricting the sample to younger age groups may lead to incorrect conclusions.

Cognitive aging may play a critical role in the development of math self-concept and math anxiety. Aging is associated with a decline in self-efficacy in cognitively demanding domains (Seeman et al., 1996), and this decline in self-efficacy can induce a decline in self-concept (Granello et al., 2025). Moreover, working memory declines during aging, making it more difficult to efficiently process complex numerical information (Salthouse, 1996). This cognitive decline may contribute to heightened anxiety when solving arithmetic problems, as older adults may struggle to maintain and manipulate numerical information in real time. While accuracy was preserved during aging, our study also found an age-related decline in reaction time across the lifespan, reflecting a slowdown in processing speed (Avcil & Artemenko, 2023). This indicates that older adults often adopt strategic adjustments to maintain performance despite cognitive slowing (Lemaire, 2024).

Notably, our exploratory analysis showed that age-related trajectories of math self-concept and math anxiety are not consistent for different levels but that developmental changes rather occur in math-anxious individuals (see Supplementary Materials 3, Figure S6). These findings underscore the importance of individual differences in shaping the developmental trajectories of math-related attitudes, recommending further research on potential moderating variables, such as educational background, career paths, and cultural influences.

Persistent association between math self-concept and math anxiety

Similar to the parallel development, a strong and stable negative correlation between math self-concept and math anxiety was observed across all age groups, suggesting that individuals with higher self confidence in math tasks show lower math anxiety levels (see also Ahmed et al., 2012). Therefore, interventions on supporting math self-concept in early education may help mitigating the increase of math anxiety (Gaspard et al., 2015).

Surprisingly, despite the strong link between math self-concept and math anxiety, their relation to math performance was not consistently found for all age groups (see also Douglas, 2000; Möller & Trautwein, 2009). As reliability was rather low in children and younger adolescents, particularly for math anxiety, we recommend using adjusted questionnaires (e.g., Carey et al., 2017) for lifespan research. After the attenuation correction for reliability, the correlations among math anxiety, math self-concept and math performance improved slightly (see Supplementary Materials 3, Table S19).

Math self-concept as a predictor of arithmetic complexity effects

Math self-concept significantly predicted arithmetic complexity effects, particularly in error rates, indicating that individuals with lower math self-concept show larger complexity effects. This corresponds to previous research showing that individuals with higher math self-concept tend to perform better in complex arithmetic tasks (Cai et al., 2018). In contrast, our results did not provide conclusive evidence for the anxiety-complexity effect (Huber & Artemenko, 2021), contrasting with prior studies suggesting that math anxiety disrupts working memory and impairs problem-solving efficiency (Ramirez et al., 2018). When math self-concept and math anxiety were both put into concurrency in predicting the carry and borrow effects, math self-concept instead of math anxiety was found to be related with the drop in performance caused by the carry and borrow operations (see also Artemenko et al., 2021).

Besides methodological limitations, a possible explanation for this discrepancy is that perceived competence (math self-concept) plays a more direct role in managing cognitive load during arithmetic processing than affective responses like anxiety. This interpretation aligns with the self-efficacy theory, which suggests that individuals with a strong belief in their abilities are more likely to persist in challenging tasks and employ effective problem-solving strategies (Bandura, 1994). Supporting this view, research has shown that discrepancies between state and trait math anxiety are closely linked to self-concept, further emphasizing its role in shaping an individual's real-time emotional and cognitive responses to math-related challenges (Roos et al., 2015).

Limitations

While this is the first lifespan study on math self-concept and math anxiety, several limitations should be acknowledged: First, the study relies on a cross-sectional design, which limits inferences on developmental changes over time and is at risk for cohort effects due to changes in school education. Second, the reliability for the questionnaire on math anxiety was relatively low in children and adolescents, recommending alternative or adjusted questionnaires for future lifespan studies. Third, the sample was drawn from the German educational system, which may limit the generalizability of the findings to other cultural or educational contexts. Math self-concept and math anxiety are known to be influenced by cultural factors, such as societal attitudes toward mathematics and gender stereotypes (Rossi et al., 2022). Future cross-cultural studies could explore how these constructs develop in different educational systems and cultural backgrounds.

Conclusions

This study provides important insights into the lifespan development of math self-concept and math anxiety, as well as their interrelation. Our findings indicate that math self-concept declines while math anxiety increases across the lifespan, with a stable negative association between these constructs. Our results highlight the particular role of aging in understanding the lifelong trajectory of math-related traits, as negative attitudes towards mathematics do not just grow out but rather intensify up to old age. This suggests that research neglecting older adults and focusing solely on younger adults may overestimate stability in adulthood. Moreover, perceived competence in math may be more influential in handling complex arithmetic tasks than affective responses such as anxiety. Hence, there is a need for early interventions to foster a positive math self-concept and reduce math anxiety, as these constructs play a crucial role in shaping mathematical engagement and performance.

References

Ahmed, W. (2018). Developmental trajectories of math anxiety during adolescence: Associations with STEM career choice. *Journal of Adolescence*, *67*, 158–166. https://doi.org/10.1016/j.adolescence.2018.06.010

Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences*, *22*(3), 385–389. https://doi.org/10.1016/j.lindif.2011.12.004

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, *109*(5), 621–634. https://doi.org/10.1037/edu0000163

Artemenko, C., Cipora, K., & Nuerk, H.-C. (2021). *Does math anxiety vary depending on situation?* https://osf.io/https://osf.io/mk39w

Artemenko, C., Soltanlou, M., Ehlis, A.-C., Nuerk, H.-C., & Dresler, T. (2018). The neural correlates of mental arithmetic in adolescents: A longitudinal fNIRS study. *Behavioral and Brain Functions*, *14*(1), 1–13.

Asendorpf, J. B. (1999). 11 Social-Personality Development. In *Individual Development from 3 to 12: Findings from the Munich Longitudinal Study* (p. 227). Cambridge University Press.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5), 181–185.

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. *Cognition & Emotion*, *8*(2), 97–125.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224. https://doi.org/10.1037/0096-3445.130.2.224

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243–248. https://doi.org/10.3758/BF03194059

Avcil, M., & Artemenko, C. (2023). *Development of arithmetic across the lifespan: A Registered Report*.

Avcil, M., & Artemenko, C. (2025). *Development of arithmetic across the lifespan: A registered Report*.

Bandura, A. (1994). *Encyclopedia of human behavior* (Vol. 4). New York: Academic Press.

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, *147*(2), 134–168. https://doi.org/10.1037/bul0000307

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860–1863.

Birgin, O., Baloğlu, M., Çatlıoğlu, H., & Gürbüz, R. (2010). An investigation of mathematics anxiety among sixth through eighth grade students in Turkey. *Learning and Individual Differences*, *20*(6), 654–658. https://doi.org/10.1016/j.lindif.2010.04.006

Cai, D., Viljaranta, J., & Georgiou, G. K. (2018). Direct and indirect effects of self-concept of ability on math skills. *Learning and Individual Differences*, *61*, 51–58. https://doi.org/10.1016/j.lindif.2017.11.009

Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The Chicken or the Egg? The Direction of the Relationship Between Mathematics Anxiety and Mathematics Performance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01987

Carey, E., Hill, F., Devine, A., & Szűcs, D. (2017). The Modified Abbreviated Math Anxiety Scale: A Valid and Reliable Instrument for Use with Children. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00011

Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics Anxiety and Science Careers among Able College Women. *Psychological Science*, *3*(5), 292–296. https://doi.org/10.1111/j.1467-9280.1992.tb00675.x

Cipora, K., Santos, F. H., Kucian, K., & Dowker, A. (2022). Mathematics anxiety-where are we and where shall we go? *Annals of the New York Academy of Sciences*, *1513*(1), 10–20. https://doi.org/10.1111/nyas.14770

Craven, E. R. G., & Marsh, H. W. (2000). *Self-Concept Theory, Research and Practice: Advances for the New Millennium*.

Ding, Y., Klapp, A., & Yang Hansen, K. (2024). The importance of mathematics self-concept and self-efficacy for mathematics achievement: A comparison between public and independent schools in Sweden. *Educational Psychology*, *44*(8), 872–892. https://doi.org/10.1080/01443410.2024.2410217

Douglas, A. F. (2000). *Math anxiety, math self-concept, and performance in math* [Thesis]. https://knowledgecommons.lakeheadu.ca/handle/2453/3140
Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, 508.
Erdogan, F., & Sengul, S. (2014). A Study on the Elementary School Students' Mathematics Self Concept. *Procedia - Social and Behavioral Sciences*, *152*, 596–601. https://doi.org/10.1016/j.sbspro.2014.09.249
Ertl, B., Luttenberger, S., & Paechter, M. (2017). The Impact of Gender Stereotypes on the Self-Concept of Female Students in STEM Subjects with an Under-Representation of Females. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00703
Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian Inference in Numerical Cognition: A Tutorial Using JASP. *Journal of Numerical Cognition*, *6*(2), 231–259. https://doi.org/10.5964/jnc.v6i2.288
Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 519–533. https://doi.org/10.1037/0012-1649.38.4.519
Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. https://doi.org/10.1037/dev0000028
Glisky, E. L. (2007). Changes in Cognitive Function in Human Aging. In *Brain Aging*. CRC Press.
Granello, F., Cuder, A., Doz, E., Pellizzoni, S., & Passolunghi, M. C. (2025). Improving math self-efficacy and math self-concept in middle school: A narrative systematic review. *European Journal of Psychology of Education*, *40*(1), 42. https://doi.org/10.1007/s10212-025-00939-5
Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124
Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics Anxiety in Young Children: An Exploratory Study. *The Journal of Experimental Education*, *81*(4), 538–555. https://doi.org/10.1080/00220973.2012.727888
Hart, S. A., & Ganley, C. M. (2019). The Nature of Math Anxiety in Adults: Prevalence and Correlates. *Journal of Numerical Cognition*, *5*(2), 122–139. https://doi.org/10.5964/jnc.v5i2.195
Helmke, A. (1999). From optimism to realism? Development of children's academic self-concept from kindergarten to grade 6. In *Individual Development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 198–221). Cambridge University Press.
Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33–46. https://doi.org/10.5951/jresematheduc.21.1.0033
Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, *48*, 45–53. https://doi.org/10.1016/j.lindif.2016.02.006
Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment*, *10*(2), 178–182.
Hopko, D. R., McNeil, D. W., Gleason, P. J., & Rabalais, A. E. (2002). The Emotional Stroop Paradigm: Performance as a Function of Stimulus Properties and Self-Reported Mathematics Anxiety. *Cognitive Therapy and Research*, *26*(2), 157–166. https://doi.org/10.1023/A:1014578218041
Huber, J. F., & Artemenko, C. (2021). Anxiety-Related Difficulties With Complex Arithmetic. *Zeitschrift Für Psychologie*.
Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008). Intraindividual Variability, Cognition, and Aging. In *The Handbook of Aging and Cognition*. Psychology Press.
Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological Research*, *71*(4), 467–483. https://doi.org/10.1007/s00426-006-0044-8
Jameson, M. M., & Fusco, B. R. (2014). Math Anxiety, Math Self-Concept, and Math Self-Efficacy in Adult Learners Compared to Traditional Undergraduate Students. *Adult Education Quarterly*, *64*(4), 306–322. https://doi.org/10.1177/0741713614541461

Klee, H. L., Miller, A. D., & Buehl, M. M. (2022). Mathematics Anxiety, Self-Concept, and Self-Efficacy: A Multidimensional Scaling Consideration of Measures. *The Journal of Experimental Education*, *91*(3), 494–516. https://doi.org/10.1080/00220973.2021.2024788
Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math Anxiety and Math Ability in Early Primary School Years. *Journal of Psychoeducational Assessment*, *27*(3), 206–225. https://doi.org/10.1177/0734282908330583
Lallement, C., & Lemaire, P. (2021). Age-related differences in how negative emotions influence arithmetic performance. *Cognition and Emotion*, *35*(7), 1382–1399. https://doi.org/10.1080/02699931.2021.1967884
Lehikoinen, H., Väisänen, P., Havu-Nuutinen, S., Lappalainen, K., & Niemivirta, M. (2024). Developmental relations between mathematics self-concept, interest, and achievement: A comparison of solo- and co-taught classes. *Instructional Science*. https://doi.org/10.1007/s11251-024-09678-4
Lemaire, P. (2024). Aging, emotion, and cognition: The role of strategies. *Journal of Experimental Psychology: General*, *153*(2), 435–453. https://doi.org/10.1037/xge0001506
Lindberg, S., Linkersdörfer, J., Ehm, J.-H., Hasselhorn, M., & Lonnemann, J. (2013). Gender Differences in Children's Math Self-Concept in the First Years of Elementary School. *Journal of Education and Learning*, *2*(3), 1–8.
Lu, Y., Li, Q., Patrick, H., & Mantzicopoulos, P. (2021). "Math Gives Me a Tummy Ache!" Mathematics Anxiety in Kindergarten. *The Journal of Experimental Education*, *89*(2), 362–378. https://doi.org/10.1080/00220973.2019.1680518
Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404–406. https://doi.org/10.1016/j.tics.2012.06.008
Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129
Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. https://doi.org/10.1037/0022-0663.81.3.417
Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology*, *82*(4), 646–656. https://doi.org/10.1037/0022-0663.82.4.646
Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*(2), 263–280. https://doi.org/10.1037/dev0000393
Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x
Marsh, H. W., & Yeung, A. S. (1997). Coursework Selection: Relations to Academic Self-Concept and Achievement. *American Educational Research Journal*, *34*(4), 691–720. https://doi.org/10.3102/00028312034004691
Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*(3), 679–690. https://doi.org/10.3758/s13428-010-0049-5
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7
Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*(1), 60–70. https://doi.org/10.1037/0022-0663.82.1.60
Möller, J., & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (pp. 179–203). Springer. https://doi.org/10.1007/978-3-540-88573-3_8
Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A Meta-Analysis of Relations Between Achievement and Self-Concept. *Review of Educational Research*, *90*(3), 376–419. https://doi.org/10.3102/0034654320919354
Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence*, *20*(2), 482–506. https://doi.org/10.1111/j.1532-7795.2010.00644.x

Namkung, J. M., Peng, P., & Lin, X. (2019). The Relation Between Mathematics Anxiety and Mathematics Performance Among School-Aged Students: A Meta-Analysis. *Review of Educational Research*, *89*(3), 459–496. https://doi.org/10.3102/0034654319843494
Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x
OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD. https://doi.org/10.1787/53f23881-en
Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9
Perinelli, E., Pisanu, F., Checchi, D., Scalas, L. F., & Fraccaroli, F. (2022). Academic self-concept change in junior high school students and relationships with academic achievement. *Contemporary Educational Psychology*, *69*, 102071. https://doi.org/10.1016/j.cedpsych.2022.102071
Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math Anxiety, Working Memory, and Math Achievement in Early Elementary School. *Journal of Cognition and Development*, *14*(2), 187–202. https://doi.org/10.1080/15248372.2012.664593
Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math Anxiety: Past Research, Promising Interventions, and a New Interpretation Framework. *Educational Psychologist*, *53*(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384
Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551–554. https://doi.org/10.1037/h0033456
Roberts, B. W., & Caspi, A. (2003). The Cumulative Continuity Model of Personality Development: Striking a Balance Between Continuity and Change in Personality Traits across the Life Course. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding Human Development: Dialogues with Lifespan Psychology* (pp. 183–214). Springer US. https://doi.org/10.1007/978-1-4615-0357-6_9
Roberts, B. W., & Mroczek, D. (2008). Personality Trait Change in Adulthood. *Current Directions in Psychological Science*, *17*(1), 31–35. https://doi.org/10.1111/j.1467-8721.2008.00543.x
Roos, A.-L., Bieg, M., Goetz, T., Frenzel, A. C., Taxer, J., & Zeidner, M. (2015). Experiencing more mathematics anxiety than expected? Contrasting trait and state anxiety in high achieving students. *High Ability Studies*, *26*(2), 245–258. https://doi.org/10.1080/13598139.2015.1095078
Rossi, S., Xenidou-Dervou, I., Simsek, E., Artemenko, C., Daroczy, G., Nuerk, H.-C., & Cipora, K. (2022). Mathematics–gender stereotype endorsement influences mathematics anxiety, self-concept, and performance differently in men and women. *Annals of the New York Academy of Sciences*, *1513*(1), 121–139. https://doi.org/10.1111/nyas.14779
Salthouse, T. A. (1996). Constraints on theories of cognitive aging. *Psychonomic Bulletin & Review*, *3*(3), 287–299. https://doi.org/10.3758/BF03210753
Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, *51*(4), 183–194. https://doi.org/10.1026/0012-1924.51.4.183
Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-Fish-Little-Pond Effect: Generalizability and Moderation—Two Sides of the Same Coin. *American Educational Research Journal*, *47*(2), 390–433. https://doi.org/10.3102/0002831209350493
Seeman, T., McAvay, G., Merrill, S., Albert, M., & Rodin, J. (1996). Self-efficacy beliefs and change in cognitive performance: MacArthur studies on Successful Aging. *Psychology and Aging*, *11*(3), 538–551. https://doi.org/10.1037/0882-7974.11.3.538
Skaalvik, S., & Skaalvik, E. M. (2005). Self-Concept, Motivational Orientation, and Help-Seeking Behavior in Mathematics: A Study of Adults Returning to High School. *Social Psychology of Education*, *8*(3), 285–302. https://doi.org/10.1007/s11218-005-3276-3
Skagerlund, K., Lind, T., Strömbäck, C., Tinghög, G., & Västfjäll, D. (2018). Financial literacy and the role of numeracy–How individuals' attitude and affinity with numbers influence financial literacy. *Journal of Behavioral and Experimental Economics*, *74*, 18–25. https://doi.org/10.1016/j.socec.2018.03.004

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

Szczygieł, M. (2019). How to measure math anxiety in young children? Psychometric properties of the modified Abbreviated Math Anxiety Scale for Elementary Children (mAMAS-E). *Polish Psychological Bulletin*, *50*(4). https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.oai-journals-pan-pl-114419/c/oai-journals-pan-pl-114419_full-text_PPB_204-19_203_20Szczygie_C5_82.pdf

Thomann, A. E., Goettel, N., Monsch, R. J., Berres, M., Jahn, T., Steiner, L. A., & Monsch, A. U. (2018). The Montreal Cognitive Assessment: Normative Data from a German-Speaking Cohort and Comparison with International Normative Samples. *Journal of Alzheimer's Disease*, *64*(2), 643–655. https://doi.org/10.3233/JAD-180080

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The Relation Between Self-Beliefs and Academic Achievement: A Meta-Analytic Review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Wortman, J., Lucas, R. E., & Donnellan, M. B. (2012). Stability and change in the Big Five personality domains: Evidence from a longitudinal study of Australians. *Psychology and Aging*, *27*(4), 867–874. https://doi.org/10.1037/a0029322

Wu, S., Amin, H., Barth, M., Malcarne, V., & Menon, V. (2012). Math Anxiety in Second and Third Graders and Its Relation to Mathematics Achievement. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00162

Yao, X., Artemenko, C., He, Y., & Nuerk, H.-C. (2025). Arithmetic is not arithmetic: Paradigm matters for arithmetic effects. *Cognition*, *256*, 106060.

Yao, X., Huber, J. F., Li, Z., Findik, Y., Nuerk, H.-C., & Artemenko, C. (2025). *The Dynamics of State Math Anxiety Vary by Paradigm and Timing During Arithmetic* (SSRN Scholarly Paper 5167231). Social Science Research Network. https://doi.org/10.2139/ssrn.5167231

A3.2 Supplementary Materials 1 for Study 3

These supplementary materials belong to the following publication:

Yao, X., Avcil, M., Meuer, P., Nuerk, H.-C., & Artemenko, C. Math self-concept decreases while math anxiety increases over the lifespan. Annals of the New York Academy of Sciences. (in revision)

Supplementary Materials 1

Results from confirmatory analysis

Title: Math self-concept decreases while math anxiety increases over the lifespan
Authors: Xinru Yao, Mine Avcil, Paul Meuer, Hans-Christoph Nuerk, Christina Artemenko

Table S1. Data exclusion on participant and trial level

	children $N = 53$	younger adolescents $N = 50$	older adolescents $N = 50$	younger adults $N = 53$	middle-aged adults $N = 50$	older adults $N = 50$
Exclusion of participants						
missing data	5	0	0	0	0	1
RT outlier in 3 MAD	1	2	0	0	3	1
ACC outlier (<25% correct)	9	1	0	0	0	2
n_excluded	15	3	0	0	3	4
Final sample	38	47	50	53	47	46
Exclusion of incomplete questionnaires						
math anxiety	2	1	0	0	0	0
math self-concept	0	0	0	0	0	1
Exclusion of trials						
incorrect response	31.62%	17.00%	10.81%	8.42%	8.73%	12.48%
RT < 200 ms	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%
RT outlier	3.31%	4.12%	3.63%	3.30%	3.97%	4.48%

Table S2. Descriptive data for math self-concept, math anxiety and math performance

		M	*SD*	*Range*
Math self-concept	children	3.30	0.61	1.75 – 4.00
	younger adolescents	3.18	0.61	1.75 – 4.00
	older adolescents	3.10	0.68	1.50 – 4.00
	younger adults	2.98	0.78	1.50 – 4.00
	middle-aged adults	2.96	0.79	1.25 – 4.00
	older adults	2.74	0.93	1.00 – 4.00
	overall	3.03	0.76	1.00 – 4.00
Math anxiety	children	1.76	0.50	1.00 – 3.33
	younger adolescents	1.94	0.50	1.00 – 3.11
	older adolescents	1.94	0.53	1.00 – 3.11
	younger adults	2.16	0.71	1.00 – 3.56
	middle-aged adults	2.19	0.90	1.00 – 4.22
	older adults	2.39	1.04	1.00 – 5.00
	overall	2.08	0.75	1.00 – 5.00
Error rate	children	0.22	0.12	0.01 – 0.44
	younger adolescents	0.16	0.11	0.00 – 0.55
	older adolescents	0.11	0.08	0.02 – 0.35
	younger adults	0.08	0.04	0.01 – 0.19
	middle-aged adults	0.09	0.06	0.00 – 0.21
	older adults	0.10	0.08	0.00 – 0.38
	overall	0.12	0.09	0.00 – 0.55
Reaction time	children	10.43	3.98	3.58 – 21.85
	younger adolescents	7.17	2.74	3.79 – 15.20
	older adolescents	5.36	2.76	1.66 – 13.05
	younger adults	4.80	1.52	2.64 – 9.42
	middle-aged adults	5.04	1.45	2.56 – 8.93
	older adults	6.31	1.78	2.25 – 9.90
	overall	6.34	3.04	1.66 – 21.85

Figure S1. Relations between math self-concept and reaction time (A) and error rate (B) in each age group

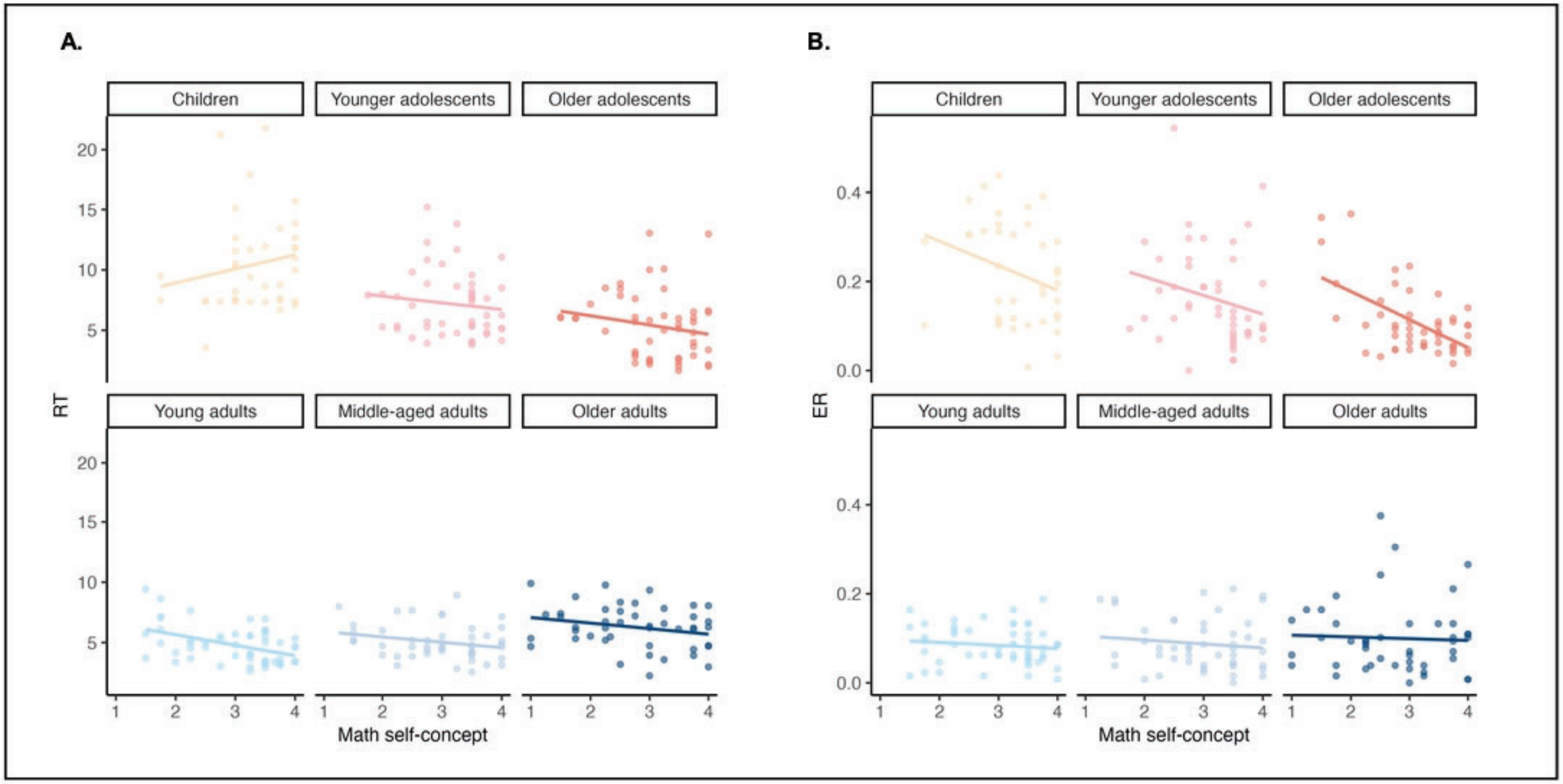


Figure S2. Relations between math anxiety and reaction time (A) and error rate (B) in each age group

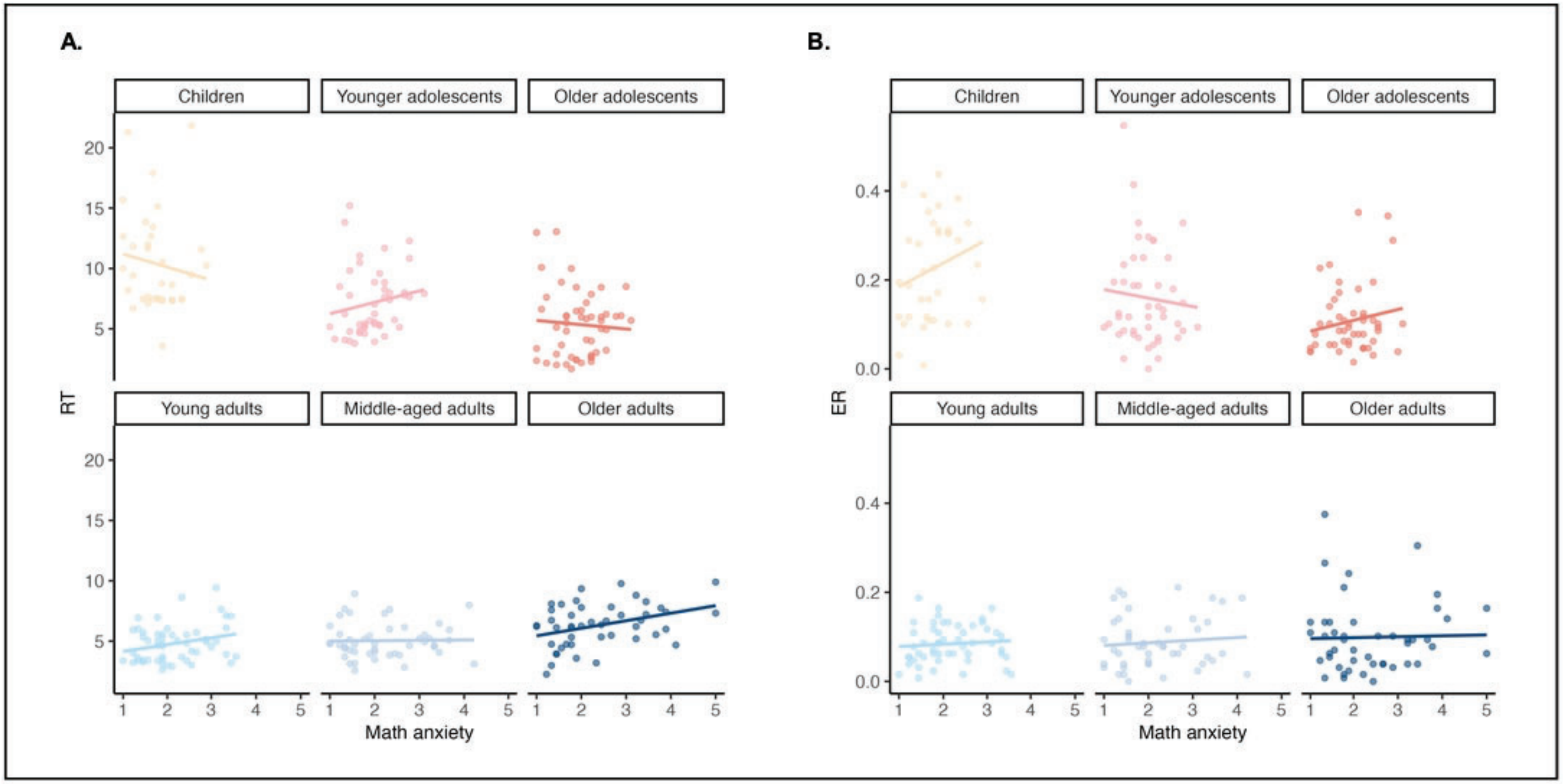


Table S3. Bayesian regression on math self-concept

(1) Model comparison for math self-concept

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	R^2
Null model	0.250	0.011	0.034	1.000	0.000
age group (linear)	0.083	0.439	8.623	> 100	0.050
age group (logarithmic)	0.083	0.251	3.693	67.224	0.036
age group (logarithmic) + age group (quadratic)	0.083	0.085	1.025	22.798	0.043
age group (linear) + age group (quadratic)	0.083	0.079	0.948	21.221	0.043
age group (linear) + age group (logarithmic)	0.083	0.077	0.917	20.587	0.043
age group (linear) + age group (logarithmic) + age group (quadratic)	0.250	0.056	0.178	4.981	0.043
age group (quadratic)	0.083	0.000	0.006	0.145	0.001

(2) Posterior summaries of coefficients

								95% Credible Interval	
Coefficient	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	$BF_{inclusion}$	*Mean*	*SD*	*Lower*	*Upper*
Intercept	1.000	0.000	1.000	0.000	1.000	3.034	0.044	2.950	3.126
age group (linear)	0.500	0.500	0.652	0.348	1.871	-0.059	0.085	-0.192	0.022
age group (logarithmic)	0.500	0.500	0.469	0.531	0.885	-0.240	0.551	-1.100	0.559
age group (quadratic)	0.500	0.500	0.221	0.779	0.284	-0.003	0.015	-0.045	0.016

Table S4. Bayesian ANOVA on math self-concept

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	BF_{01}
Null model	0.500	0.361	0.566	1.000	
age group	0.500	0.639	1.766	1.766	0.566

(2) Analysis of effects

Effects	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	BF_{incl}	BF_{excl}
age group	0.500	0.500	0.639	0.361	1.766	0.566

Table S5. Bayesian regression on math anxiety

(1) Model comparison for math anxiety

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	R^2
Null model	0.250	0.001	0.004	1.000	0.000
age group (linear)	0.083	0.484	10.300	> 100	0.066
age group (logarithmic)	0.083	0.217	3.055	> 100	0.060
age group (linear) + age group (quadratic)	0.083	0.083	0.997	> 100	0.066
age group (linear) + age group (logarithmic)	0.083	0.082	0.984	> 100	0.066
age group (logarithmic) + age group (quadratic)	0.083	0.080	0.951	> 100	0.066
age group (linear) + age group (logarithmic) + age group (quadratic)	0.250	0.053	0.168	40.016	0.066
age group (quadratic)	0.083	0.000	0.000	0.152	0.001

(2) Posterior summaries of coefficients

								95% Credible Interval	
Coefficient	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	$BF_{inclusion}$	*Mean*	*SD*	*Lower*	*Upper*
Intercept	1.000	0.000	1.000	0.000	1.000	2.077	0.044	2.000	2.163
age group (linear)	0.500	0.500	0.702	0.298	2.352	0.079	0.086	0.000	0.211
age group (logarithmic)	0.500	0.500	0.432	0.568	0.760	0.214	0.553	-0.511	1.112
age group (quadratic)	0.500	0.500	0.216	0.784	0.275	0.003	0.014	-0.011	0.035

Table S6. Bayesian ANOVA on math anxiety

(1) Model comparisons

Models	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	BF_{01}
Null model	0.500	0.190	0.081	1.000	
age group	0.500	0.810	12.404	12.404	0.081

(2) Analysis of effects

Effects	$P(incl)$	$P(excl)$	$P(incl\|data)$	$P(excl\|data)$	BF_{incl}	BF_{excl}
age group	0.500	0.500	0.925	0.075	12.404	0.081

(3) Post hoc comparisons

		Prior Odds	Posterior Odds	$BF_{10,\ U}$	error %
children	younger adults	0.260	2.219	8.535	0.000
	middle-aged adults	0.260	1.066	4.102	0.007
	older adults	0.260	6.471	24.895	0.000
older adults	younger adolescents	0.260	1.155	4.442	0.007
	older adolescents	0.260	1.250	4.809	0.006

Table S7. Bayesian ANCOVA on error rate

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	*error %*
Null model	0.077	0.000	0.000	1.000	
age group + math self-concept	0.077	0.665	23.771	> 100	0.839
age group + math self-concept + math anxiety	0.077	0.172	2.499	> 100	0.871
age group	0.077	0.120	1.636	> 100	0.001
age group + math anxiety	0.077	0.035	0.441	> 100	1.083
age group + math self-concept + age group × math self-concept	0.077	0.005	0.059	> 100	1.049
age group + math self-concept + math anxiety + age group × math self-concept	0.077	0.002	0.021	> 100	5.112
age group + math self-concept + math anxiety + age group × math anxiety	0.077	0.000	0.009	> 100	5.912
age group + math anxiety + age group × math anxiety	0.077	0.000	0.002	> 100	5.287
age group + math self-concept + math anxiety + age group × math self-concept + age group × math anxiety	0.077	0.000	0.000	> 100	11.141

(2) Analysis of effects

Effects	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	BF_{incl}	BF_{excl}
age group	0.308	0.308	0.992	0.000	> 100	0.000
math self-concept	0.385	0.385	0.838	0.156	5.385	0.185
math anxiety	0.385	0.385	0.210	0.789	0.265	3.853
age group × math self-concept	0.231	0.231	0.007	0.838	0.008	> 100
age group × math anxiety	0.231	0.231	9.829×10^{-4}	0.210	0.005	> 100

Table S8. Bayesian ANCOVA on reaction time

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	*error %*
Null model	0.077	0.000	0.000	1.000	
math self-concept + age group	0.077	0.453	9.927	> 100	0.764
age group	0.077	0.318	5.588	> 100	0.004
math anxiety + age group	0.077	0.103	1.380	> 100	1.382
math self-concept + math anxiety + age group	0.077	0.092	1.222	> 100	1.077
math self-concept + age group + math self-concept × age group	0.077	0.017	0.210	> 100	4.392
math self-concept + math anxiety + age group + math self-concept × age group	0.077	0.006	0.074	> 100	5.983
math self-concept + math anxiety + age group + math anxiety × age group	0.077	0.006	0.068	> 100	5.396
math anxiety + age group + math anxiety × age group	0.077	0.004	0.043	> 100	5.323
math self-concept + math anxiety + age group + math self-concept × age group + math anxiety × age group	0.077	0.002	0.018	> 100	8.943

(2) Analysis of effects

Effects	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	BF_{incl}	BF_{excl}
age group	0.308	0.308	0.966	0.000	> 100	0.000
math self-concept	0.385	0.385	0.551	0.424	1.298	0.754
math anxiety	0.385	0.385	0.202	0.788	0.256	3.945
age group × math self-concept	0.231	0.231	0.025	0.551	0.045	22.850
age group × math anxiety	0.231	0.231	0.011	0.202	0.053	18.878

Table S9. Bayesian ANCOVA on complexity effect (error rate)

(1) Model comparison

Models	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	$error\ \%$
Null model	0.077	0.009	0.105	1.000	
age group + math self-concept	0.077	0.622	19.765	71.906	1.114
math self-concept	0.077	0.147	2.064	16.958	0.004
age group + math self-concept + math anxiety	0.077	0.132	1.819	15.211	0.513
math self-concept + math anxiety	0.077	0.046	0.574	5.276	0.006
age group + math self-concept + math anxiety + age group × math anxiety	0.077	0.014	0.167	1.587	5.261
age group + math self-concept + age group × math self-concept	0.077	0.013	0.154	1.464	4.249
age group + math self-concept + math anxiety + age group × math self-concept	0.077	0.006	0.075	0.714	4.948
age group + math anxiety	0.077	0.005	0.057	0.549	0.894
age group	0.077	0.004	0.050	0.477	0.000

(2) Analysis of effects

Effects	$P(incl)$	$P(excl)$	$P(incl\|data)$	$P(excl\|data)$	BF_{incl}	BF_{excl}
age group	0.308	0.308	0.763	0.204	3.739	0.258
math self-concept	0.385	0.385	0.960	0.021	45.991	0.021
math anxiety	0.385	0.385	0.191	0.794	0.241	4.331
age group × math self-concept	0.231	0.231	0.019	0.768	0.025	41.920
age group × math anxiety	0.231	0.231	0.014	0.143	0.101	10.642

Table S10. Bayesian regression on complexity effect (error rate)

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	R^2
Null model	0.278	0.002	0.004	1.000	0.000
age group (linear) + math self-concept	0.028	0.517	37.511	> 100	0.087
age group (linear) + math self-concept + age group (linear) × math self-concept	0.028	0.125	4.992	> 100	0.088
age group (linear) + math self-concept + math anxiety	0.028	0.115	4.568	> 100	0.088
age group (linear) + math self-concept + math anxiety + age group (linear) × math self-concept + age group (linear) × math anxiety	0.278	0.102	0.297	63.676	0.090
age group (linear) + math self-concept + math anxiety + age group (linear) × math self-concept	0.056	0.070	1.285	> 100	0.090
age group (linear) + math self-concept + math anxiety + age group (linear) × math anxiety	0.056	0.055	0.981	> 100	0.088
math self-concept	0.056	0.005	0.080	14.612	0.035
age group (linear) + math anxiety	0.028	0.004	0.131	23.185	0.051
age group (linear)	0.056	0.003	0.058	10.585	0.033

(2) Posterior summaries of coefficients

Coefficient	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	$BF_{inclusion}$	*Mean*	*SD*	95% Credible Interval *Lower*	95% Credible Interval *Upper*
Intercept	1.000	0.000	1.000	0.000	1.000	0.254	0.025	0.206	0.298
age group (linear)	0.139	0.417	0.640	0.007	> 100	-0.039	0.060	-0.122	0.115
math self-concept	0.194	0.444	0.693	0.010	> 100	-0.128	0.039	-0.204	-0.058
math anxiety	0.194	0.444	0.190	0.652	0.667	-0.007	0.030	-0.095	0.040
age group (linear) × math self-concept	0.361	0.111	0.298	0.687	0.133	-0.005	0.014	-0.040	0.020
age group (linear) × math anxiety	0.361	0.111	0.158	0.189	0.256	-0.001	0.011	-0.028	0.017

Table S11. Bayesian ANCOVA on complexity effect (reaction time)

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	*error %*
Null model	0.077	0.000	0.000	1.000	
age group	0.077	0.663	23.647	> 100	0.000
math self-concept + age group	0.077	0.132	1.819	> 100	0.942
math anxiety + age group	0.077	0.117	1.595	> 100	1.818
math anxiety + age group + math anxiety × age group	0.077	0.038	0.473	> 100	4.267
math self-concept + math anxiety + age group	0.077	0.027	0.338	> 100	1.026
math self-concept + math anxiety + age group + math anxiety × age group	0.077	0.016	0.199	> 100	5.129
math self-concept + age group + math self-concept × age group	0.077	0.004	0.045	> 100	4.790
math self-concept + math anxiety + age group + math self-concept × age group	0.077	0.001	0.015	> 100	6.783
math self-concept + math anxiety + age group + math self-concept × age group + math anxiety × age group	0.077	0.001	0.014	> 100	8.967

(2) Analysis of effects

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF_{incl}	BF_{excl}
age group	0.308	0.308	0.940	0.000	> 100	0.000
math self-concept	0.385	0.385	0.175	0.819	0.214	4.550
math anxiety	0.385	0.385	0.146	0.799	0.183	5.624
age group × math self-concept	0.231	0.231	0.006	0.175	0.035	30.930
age group × math anxiety	0.231	0.231	0.055	0.146	0.379	2.759

Table S12. Bayesian regression on complexity effect (reaction time)

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	R^2
Null model	0.278	0.000	0.000	1.000	0.000
age group (linear)	0.056	0.553	21.014	> 100	0.116
age group (linear) + math self-concept + age group (linear) × math self-concept	0.028	0.096	3.705	> 100	0.133
age group (linear) + math self-concept + math anxiety + age group (linear) × math self-concept + age group (linear) × math anxiety	0.278	0.070	0.197	> 100	0.135
age group (linear) + math anxiety + age group (linear) × math anxiety	0.028	0.067	2.528	> 100	0.130
age group (linear) + math self-concept	0.028	0.067	2.522	> 100	0.119
age group (linear) + math anxiety	0.028	0.056	2.085	> 100	0.118
age group (linear) + math self-concept + math anxiety + age group (linear) × math self-concept	0.056	0.043	0.759	> 100	0.133
age group (linear) + math self-concept + math anxiety + age group (linear) × math anxiety	0.056	0.035	0.614	> 100	0.132
age group (linear) + math self-concept + math anxiety	0.028	0.013	0.452	> 100	0.119

(2) Posterior summaries of coefficients

Coefficient	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	$BF_{inclusion}$	*Mean*	*SD*	95% Credible Interval *Lower*	95% Credible Interval *Upper*
Intercept	1.000	0.000	1.000	0.000	1.000	1808.089	75.088	1666.404	1945.101
age group (linear)	0.139	0.417	0.689	0.000	> 100	-232.876	209.301	-584.967	300.552
math self-concept	0.194	0.444	0.115	0.676	0.388	-19.778	74.677	-237.859	81.462
math anxiety	0.194	0.444	0.112	0.716	0.357	-9.721	79.139	-211.447	150.781
age group (linear) × math self-concept	0.361	0.111	0.209	0.115	0.559	-22.704	54.712	-162.157	0.000
age group (linear) × math anxiety	0.361	0.111	0.173	0.112	0.475	17.241	49.611	0.000	150.915

A3.3 Supplementary Materials 2 for Study 3

These supplementary materials belong to the following publication:

Yao, X., Avcil, M., Meuer, P., Nuerk, H.-C., & Artemenko, C. Math self-concept decreases while math anxiety increases over the lifespan. Annals of the New York Academy of Sciences. (in revision)

Supplementary Materials 2

Results from exploratory analysis on gender

Title: Math self-concept decreases while math anxiety increases over the lifespan
Authors: Xinru Yao, Mine Avcil, Paul Meuer, Hans-Christoph Nuerk, Christina Artemenko

Exploratory analyses on gender differences

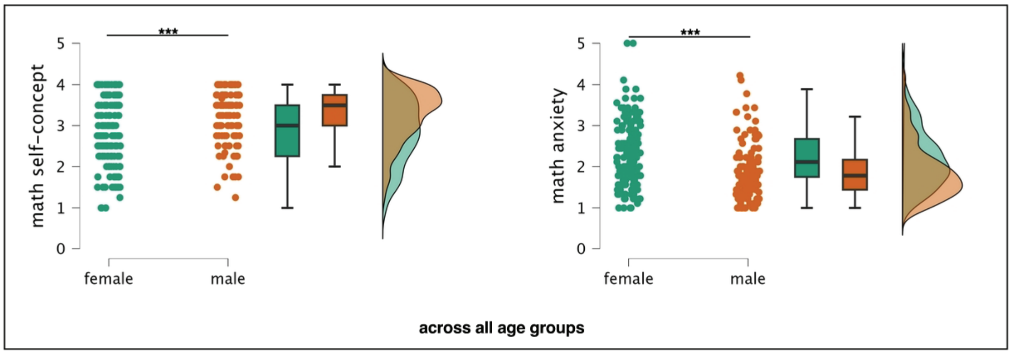
Bayesian independent samples *t*-test revealed that females reported a lower math self-concept (BF_{10} > 100) but higher math anxiety (BF_{10} > 100) than males across all age groups (see Figure S3). However, Bayesian independent samples *t*-test for each age group were not consistent with evidence for a lower math self-concept in females than in males (see Figure S4) only in children (BF_{10} = 9.81) and older adults (BF_{10} = 35.65), and evidence for higher math anxiety in females than in males (see Figure S5) only in children (BF_{10} = 26.89), older adolescents (BF_{10} = 15.90), and older adults (BF_{10} = 7.68).

Table S13. Bayesian *t*-tests for gender differences in math self-concept and math anxiety

	children		younger adolescents		older adolescents		younger adults		middle-aged adults		older adults	
	BF_{10}	BF_{01}	BF_{10}	BF_{01}	BF_{10}	BF_{01}	BF_{10}	BF_{01}	BF_{10}	BF_{01}	BF_{10}	BF_{01}
math self-concept	**9.81**	0.10	0.63	1.59	1.17	0.85	0.345	2.90	0.47	2.11	**35.65**	0.03
math anxiety	**26.89**	0.04	0.53	1.89	**15.90**	0.06	0.33	3.06	0.31	3.23	**7.68**	0.13

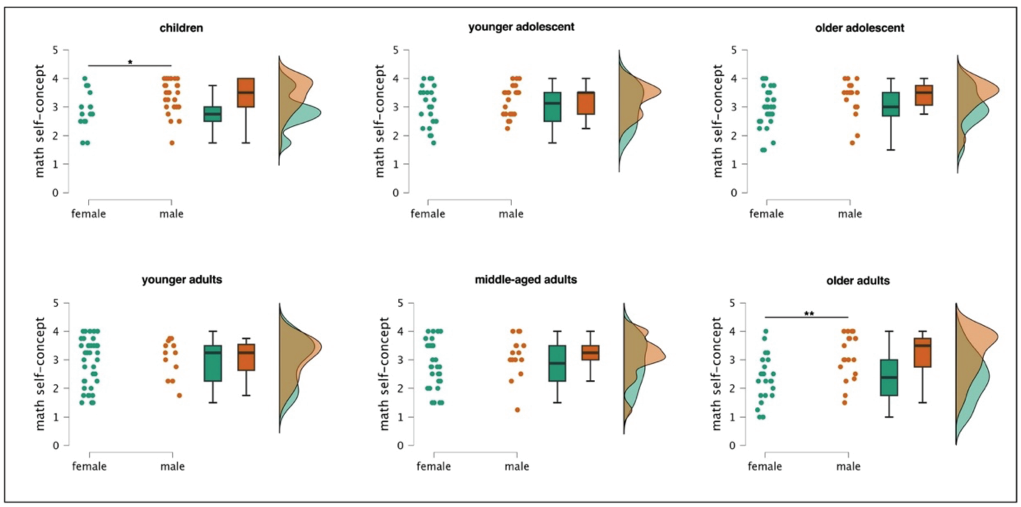
Due to the observed gender differences, the confirmatory analyses were repeated as exploratory analyses including gender. The results on math self-concept only changed in the regard that there was conclusive (instead of inconclusive) evidence for the linear predictor of age group after controlling for gender (see Table S14) and no changes in the ANOVA (see Table S15). The results on math anxiety only changed in the regard that there was conclusive (instead of inconclusive) evidence for the linear predictor of age group after controlling for gender in the regression (see Table S16) and vice versa in the ANOVA (see Table S17). The results for the analyses on math performance and the complexity effect did not change.

Figure S3. Gender differences across all age groups



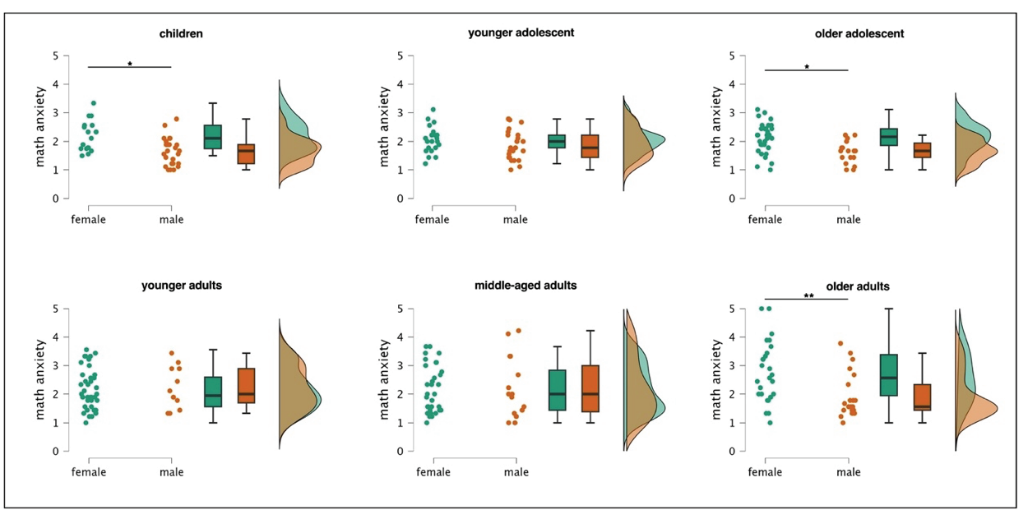
Notes. *BF_{10} > 3, **BF_{10} > 30, ***BF_{10} > 100

Figure S4. Gender comparison for math self-concept in each age group



Notes. *BF_{10} > 3, **BF_{10} > 30, ***BF_{10} > 100

Figure S5. Gender comparison for math anxiety in each age group



Notes. *BF_{10} > 3, **BF_{10} > 30, ***BF_{10} > 100

Table S14. Bayesian regression on math self-concept (controlling for gender)

(1) Model comparison for math self-concept

Models	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	R^2
Null model	0.252	0.000	0.000	1.000	0.000
age group (linear) + gender	0.012	0.134	12.783	> 100	0.111
age group (logarithmic) + age group (quadratic) + gender + age group (quadratic) × gender	0.007	0.111	17.216	> 100	0.135
age group (linear) + age group (quadratic) + gender + age group (quadratic) × gender	0.007	0.110	17.138	> 100	0.135
age group (linear) + age group (logarithmic) + age group (quadratic) + gender + age group (linear) × gender + age group (logarithmic) × gender + age group (quadratic) × gender	0.252	0.106	0.352	> 100	0.138
age group (logarithmic) + gender	0.012	0.051	4.447	> 100	0.105
age group (logarithmic) + age group (quadratic) + gender + age group (logarithmic) × gender + age group (quadratic) × gender	0.012	0.051	4.419	> 100	0.136
age group (linear)+ age group (quadratic) + gender + age group (linear) × gender + age group (quadratic) × gender	0.012	0.048	4.117	> 100	0.136
age group (linear) + age group (logarithmic) + age group (quadratic) + gender + age group (quadratic) × gender	0.012	0.044	3.795	> 100	0.135
age group (linear) + age group (logarithmic) + age group (quadratic) + gender + age group (logarithmic) × gender + age group (quadratic) × gender	0.036	0.040	1.106	> 100	0.136

(2) Posterior summaries of coefficients

								95% Credible Interval	
Coefficient	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	$BF_{inclusion}$	*Mean*	*SD*	*Lower*	*Upper*
Intercept	1.000	0.000	1.000	0.000	1.000	3.032	0.043	2.944	3.123
age group (linear)	0.175	0.444	0.404	0.334	3.064	-0.052	0.189	-0.406	0.452
age group (logarithmic)	0.175	0.444	0.326	0.425	1.944	-0.045	1.231	-2.974	2.683
age group (quadratic)	0.175	0.444	0.128	0.297	1.093	0.009	0.042	-0.067	0.102
gender	0.101	0.403	0.329	0.000	> 100	-0.108	0.867	-1.769	1.794
age group (linear) × gender	0.381	0.108	0.262	0.403	0.184	0.029	0.225	-0.454	0.504
age group (logarithmic) × gender	0.381	0.108	0.248	0.326	0.216	-0.260	1.502	-3.792	2.373
age group (quadratic) × gender	0.381	0.108	0.575	0.128	1.271	-0.054	0.063	-0.165	0.011

Table S15. Bayesian ANOVA on math self-concept (controlling for gender)

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	BF_{01}
Null model (incl. gender)	0.333	0.630	3.407	1.000	1.000
age group	0.333	0.286	0.800	0.453	1.228
age group + age group × gender	0.333	0.084	0.184	0.134	1.233

(2) Analysis of effects

Effects	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	BF_{incl}	BF_{excl}
age group	0.333	0.333	0.286	0.630	0.453	2.216
age group × gender	0.333	0.333	0.084	0.286	0.295	3.438

Table S16. Bayesian regression on math anxiety (controlling for gender)

(1) Model comparison for math anxiety

Models	$P(M)$	$P(M\|data)$	BF_M	BF_{10}	R^2
Null model	0.252	0.000	0.000	1.000	0.000
age group (linear) + age group (logarithmic) + age group (quadratic) + gender + age group (linear) × gender + age group (logarithmic) × gender + age group (quadratic) × gender	0.252	0.233	0.902	> 100	0.127
age group (linear) + gender	0.012	0.201	20.717	> 100	0.098
age group (logarithmic) + gender	0.012	0.063	5.529	> 100	0.090
age group (linear) + age group (quadratic) + gender + age group (quadratic) × gender	0.007	0.059	8.667	> 100	0.115
age group (logarithmic) + age group (quadratic) + gender + age group (quadratic) × gender	0.007	0.054	7.872	> 100	0.114
age group (logarithmic) + age group (quadratic) + gender	0.007	0.038	5.525	> 100	0.101
age group (linear) + age group (quadratic) + gender	0.007	0.038	5.396	> 100	0.101
age group (linear)	0.036	0.034	0.930	> 100	0.065
age group (linear) + age group (logarithmic) + gender	0.007	0.032	4.605	> 100	0.100

(2) Posterior summaries of coefficients

								95% Credible Interval	
Coefficient	$P(incl)$	$P(excl)$	$P(incl\|data)$	$P(excl\|data)$	$BF_{inclusion}$	$Mean$	SD	$Lower$	$Upper$
Intercept	1.000	0.000	1.000	0.000	1.000	2.079	0.043	1.983	2.150
age group (linear)	0.175	0.444	0.434	0.223	4.928	0.204	0.329	-0.211	1.106
age group (logarithmic)	0.175	0.444	0.273	0.405	1.709	-0.783	2.176	-6.851	1.395
age group (quadratic)	0.175	0.444	0.123	0.412	0.755	-0.029	0.066	-0.194	0.054
gender	0.101	0.403	0.387	0.055	28.215	-0.647	1.808	-5.277	0.484

								95% Credible Interval	
Coefficient	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	$BF_{inclusion}$	*Mean*	*SD*	*Lower*	*Upper*
age group (linear) × gender	0.381	0.108	0.342	0.397	0.244	-0.214	0.473	-1.451	0.103
age group (logarithmic) × gender	0.381	0.108	0.321	0.254	0.357	1.468	3.159	-0.430	9.071
age group (quadratic) × gender	0.381	0.108	0.465	0.119	1.109	0.068	0.101	0.000	0.292

Table S17. Bayesian ANOVA on math anxiety (controlling for gender)

(1) Model comparison

Models	*P(M)*	*P(M|data)*	BF_M	BF_{10}	BF_{01}
Null model	0.200	0.001	0.005	1.000	
age group + gender	0.200	0.456	3.350	> 100	
age group + gender + age group × gender	0.200	0.349	2.143	> 100	
gender	0.200	0.180	0.880	> 100	
age group	0.200	0.014	0.056	10.650	0.394

(2) Analysis of effects

Effects	*P(incl)*	*P(excl)*	*P(incl|data)*	*P(excl|data)*	BF_{incl}	BF_{excl}
age group	0.400	0.400	0.469	0.182	2.584	0.394
gender	0.400	0.400	0.636	0.015	42.478	
age group × gender	0.200	0.200	0.349	0.456	0.765	1.253

A3.4 Supplementary Materials 3 for Study 3

These supplementary materials belong to the following publication:

Yao, X., Avcil, M., Meuer, P., Nuerk, H.-C., & Artemenko, C. Math self-concept decreases while math anxiety increases over the lifespan. Annals of the New York Academy of Sciences. (in revision)

Supplementary Materials 3

Results from additional exploratory analysis

Title: Math self-concept decreases while math anxiety increases over the lifespan
Authors: Xinru Yao, Mine Avcil, Paul Meuer, Hans-Christoph Nuerk, Christina Artemenko

Exploratory analysis of efficiency scores

To explore whether the inconsistent pattern of correlations between math self-concept, anxiety and performance is due to speed-accuracy trade-offs in the different age groups, efficiency scores were calculated to consider both reaction time and accuracy:

Efficiency score = Median reaction time / Percentage of correct answers

Efficiency scores are a composite measure used to assess overall task performance by accounting for both speed and accuracy. A higher efficiency score indicates lower overall efficiency and task performance (longer reaction times and/or lower accuracy). Conversely, a lower efficiency score reflects better task performance (faster and/or more accurate responses).

To explore whether math self-concept and math anxiety are correlated with efficiency scores within or across age groups, Bayesian correlations were evaluated for each age group and across groups. Results showed conclusive evidence only for the correlation between math self-concept and efficiency score in younger adults (see Table S18), indicating that higher math self-concept is related to better task performance in younger adults.

Table S18. Correlations of efficiency scores within and across age groups

	children	younger adolescents	older adolescents	younger adults	middle-aged adults	older adults	overall
math self-concept	.013	-.166	-.308	-.436**	-.235	-.149	-.031
math anxiety	-.082	.086	-.001	.227	.037	.207	-.063

Notes. * $BF_{10} > 3$, ** $BF_{10} > 30$, *** $BF_{10} > 100$

Exploratory analysis for different levels of math self-concept and math anxiety

To further investigate the developmental changes in math self-concept and math anxiety across the lifespan, we categorized both variables into three levels based on the whole sample: low (-1 *SD*), average, and high (+1 *SD*). This categorization allowed us to examine whether the linear age-related trends differ across different levels of math self-concept and math anxiety, by accounting for individual differences across age groups. We conducted Bayesian regressions on math self-concept and math anxiety by considering age group and math self-concept/anxiety levels, and their interaction.

Math self-concept

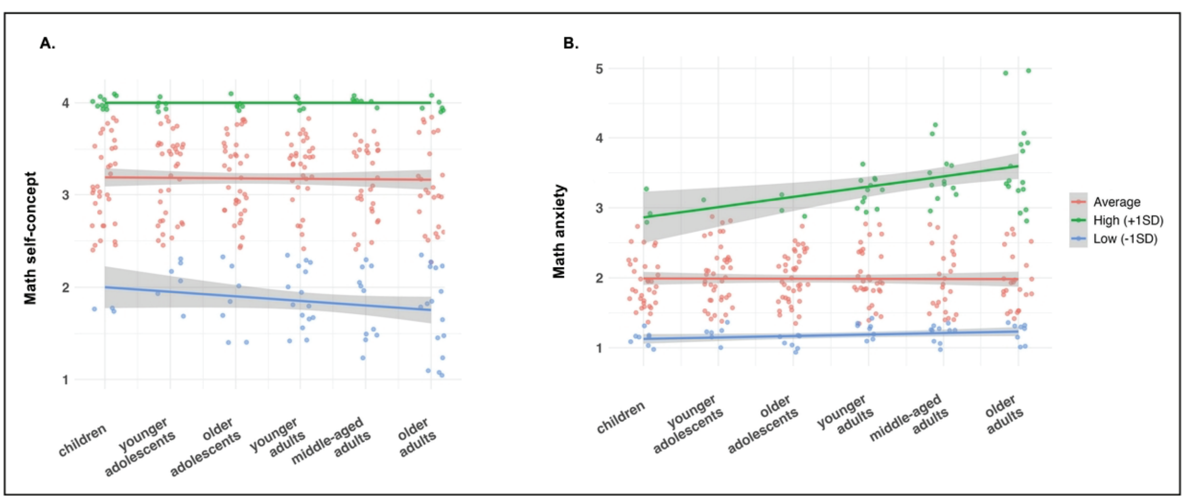
The exploratory analysis on math self-concept revealed inconclusive evidence for age group (β = -0.01, 95% CI [-0.06, 0.04]) and for the interaction between age and math self-concept levels, as neither the high (+1 SD) group (β = 0.00, 95% CI [-0.07, 0.08]) nor the low (-1 SD) group (β = -0.04, 95% CI [-0.11, 0.04]) showed significant changes over time. This suggests that the relationship between math self-concept and age remains relatively stable across different self-concept levels, with no clear evidence that age moderates this association.

Math anxiety

The exploratory analysis on math anxiety showed inconclusive evidence for age (β = -0.00, 95% CI [-0.04, 0.04]), but conclusive evidence for the interaction between age and high math anxiety (β = 0.15, 95% CI [0.07, 0.23]), indicating that individuals with high anxiety exhibited showed even higher levels of math anxiety in older than younger individuals. This suggests that math anxiety levels of highly anxious individuals are more pronounced with increasing age, potentially due to cumulative negative experiences or avoidance behaviors in mathematical contexts. Conversely, there was inconclusive evidence for the interaction between age and low anxiety (β = 0.02, 95% CI [-0.05, 0.09]).

These findings suggest that the age-related trajectories of math self-concept and math anxiety are not consistent for different levels (see Figure S6): Math anxiety, particularly for highly math-anxious individuals, is more pronounced with aging. While the descriptive pattern is similar for math self-concept, evidence rather revealed that it remains relatively stable over time.

Figure S6. Different levels of math self-concept and math anxiety across the lifespan



Exploratory analysis with attenuation correction for reliability of math anxiety

S19. Corrected Bayesian Correlation between Math anxiety and reaction time

	children	younger adolescents	older adolescents	younger adults	middle-aged adults	older adults	overall
math self-concept & math anxiety	-0.813 **	-0.573 *	-0.760 ***	-0.796 ***	-0.791 ***	-0.826 ***	-0.786 ***
math self-concept & reaction time	0.210	-0.136	-0.200	-0.466 **	-0.253	-0.251	-0.025
math self-concept & error rate	-0.305	-0.180	-0.523 **	-0.085	-0.126	-0.124	-0.088
math anxiety & reaction time	-0.163	0.201	-0.077	0.280	0.019	0.381 *	-0.028
math anxiety & error rate	0.311	-0.066	0.160	0.101	0.094	0.100	-0.014
math self-concept & complexity effect (RT)	0.235	-0.015	-0.016	-0.382 *	-0.230	-0.047	-0.088
math self-concept & complexity effect (ER)	0.046	-0.294	-0.025	-0.371 *	-0.363	-0.184	-0.191
math anxiety & complexity effect (RT)	-0.191	0.287	-0.266	0.339	0.065	0.115	-0.051
math anxiety& complexity effect (ER)	0.040	-0.008	-0.154	0.402 *	0.240	0.056	0.091

A4. Study 4

A4.1 Publication of Study 4

Yao, X., Barth, B., & Artemenko, C. When arithmetic gets complex: fNIRS evidence for fronto-parietal activation in multi-digit arithmetic. *npj* Science of Learning (under review)

When arithmetic gets complex: fNIRS evidence for fronto-parietal activation in multi-digit arithmetic

Xinru Yao[1, 2 *], Beatrix Barth[2,3,4], Christina Artemenko[1, 2, 4*]

[1] Department of Psychology, University of Tuebingen, Tuebingen, Germany

[2] LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany

[3] Department of Psychiatry and Psychotherapy, Tuebingen Center for Mental Health, University Hospital of Tuebingen, Tuebingen, Germany

[4] German Center for Mental Health (DZPG), partner site Tuebingen, Germany

* Correspondence concerning this article should be addressed to Xinru Yao:

Name: Xinru Yao

Email address: xinru.yao.tuebingen@gmail.com

Abstract

Arithmetic is represented in a fronto-parietal network. When the complexity in multi-digit arithmetic increases, the question arises whether primarily domain-general processing demands are increased, such as working memory in the frontal cortex, or also domain-specific processing demands, such as magnitude and place-value processing in the parietal cortex. This study addressed this question using functional near-infrared spectroscopy (fNIRS) to evaluate the neural correlates of multi-digit arithmetic in a written production paradigm. During fNIRS, adults (behavior: $N = 48$; fNIRS: $N = 46$) solved three-digit addition and subtraction problems with no, 1, or 2 carry or borrow operations, reflecting task complexity. At the behavioral level, complexity was reflected by reduced performance, i.e., lower accuracy and longer reaction times. At the neural level, fNIRS results revealed main effects of complexity in the bilateral fronto-parietal network of arithmetic processing, particularly in the bilateral inferior frontal gyrus (IFG), middle frontal gyrus (MFG), intraparietal sulcus (IPS), angular gyrus (AG), and supramarginal gyrus (SMG). Working memory differentially buffered complexity effects: larger verbal memory was linked to smaller accuracy costs and less frontal recruitment, whereas larger visuospatial memory was linked to less parietal recruitment. Together, these results show that the difficulty of multi-digit arithmetic relies on the bilateral fronto-parietal network, reflecting the joint contribution of domain-general and domain-specific processes.

Keywords:

multi-digit arithmetic, carry effect, borrow effect, fNIRS, complexity, working memory, fronto-parietal network

Introduction

Multi-digit arithmetic is a cognitively demanding skill that plays a crucial role in both educational and everyday contexts. It requires the integration of multiple processes, including working memory, place-value understanding, and carry/borrow procedures. Despite its importance, our understanding of how the brain manages these complex operations remains limited. In the field of educational neuroscience, bridging this knowledge gap is essential, as it can offer insights that impact mathematics education. As AI-driven tutoring and assessment become increasingly common, understanding the neural basis of human arithmetic can inform the design of intelligent learning systems and provide benchmarks for evaluating AI performance against human cognition. Traditionally, neuroimaging studies have predominantly focused on single-digit arithmetic and employed simplified yes/no verification paradigms due to methodological constraints, which may deviate from brain activation during arithmetic tasks in natural settings (Uittenhove et al., 2015; Yao et al., 2025). Functional near-infrared spectroscopy (fNIRS) enables the implementation of production paradigms in arithmetic research, as it is relatively motion-insensitive, applicable in upright positions and naturalistic settings (Soltanlou et al., 2018).

Arithmetic relies on a fronto-parietal network (Istomina & Arsalidou, 2024) that integrates domain-general cognitive processes and domain-specific numerical processes. Within the two-network framework 11/11/2025 23:38:00, this fronto-parietal network is associated with magnitude processing, centred around the intraparietal sulcus (IPS) and connected to prefrontal areas. The prefrontal areas include the inferior frontal gyrus (IFG) and middle frontal gyrus (MFG), which are mainly associated with domain-general processes such as working memory and executive control used to temporarily store and manipulate intermediate results during multi-step calculations. On the other hand, the IPS is primarily associated with domain-specific processes, such as number magnitude and place-value processing (Arsalidou & Taylor, 2011; Emerson & Cantlon, 2012; Park & Brannon, 2014).

In two-digit arithmetic, complexity increases when carry or borrow operations are necessary and thus computations across place-values (Nuerk et al., 2015): The carry operation in addition is required when the sum of the units of operands exceeds 9, with a decade to be carried over (e.g., 36 + 27) and, similarly, the borrow operation in subtraction is required when the unit of the subtrahend is larger than the unit of minuend and hence a decade has to be borrowed (e.g., 63 – 27). Behaviorally, the carry effect in addition and the borrow effect in subtraction increase arithmetic difficulty in terms of accuracy and response times (Artemenko, 2018; Imbo & LeFevre, 2010; Moeller et al., 2011) and reflect the demands of place-value computation (Nuerk et al., 2015). However, previous findings based on single- and two-digit arithmetic (Kong et al., 2005; Yi-Rong et al., 2011) cannot be directly generalized to multi-digit arithmetic, as behavioral studies suggest that different strategies and cognitive demands are involved (Bahnmueller et al., 2015; Meyerhoff et al., 2012). The cognitive demands associated with carry and borrow operations increase with the number of place-value transformations (Imbo et al., 2005). This is evident in three-digit arithmetic: a single problem may involve no, 1 or 2 carry/borrow operations (e.g., 425 + 512 needs no carry, 548 + 236 needs 1 carry, 152 + 569 needs 2 carries), and each additional operation requires an extra place-value shift and storage of intermediate digits (Imbo et al., 2007b; Selter, 2001). Therefore, explicitly manipulating the number of carry/borrow operations provides test of how domain-specific place-value processing and domain-general working memory demands jointly contribute to arithmetic complexity.

Neuroimaging evidence indicates that carry and borrow operations elicit activation within the fronto-parietal network of arithmetic processing. Frontal regions are particularly sensitive to increased task demands due to arithmetic complexity. Increased activation in the left IFG and MFG has been observed during problems requiring carry or borrow, reflecting elevated demands on working memory and cognitive control (Verner et al., 2013). At the same time, there is some evidence for the involvement of the IPS in carry and borrow operations, consistent with increased numerical magnitude and place-value processing demands (Artemenko et al., 2015; Klein et al., 2010). These mechanisms are expected to play an even greater role in three-digit arithmetic, where multiple carry and borrow operations can occur within a single problem (Butterworth et al., 2001; Imbo et al., 2007a). However, no neuroimaging study to date has examined how these regions contribute to multiple carry or borrow steps – operations that substantially increase task complexity in multi-digit arithmetic. The question is whether processing this increased complexity exclusively relies on frontal resources or also affects IPS activation (even if problem size is controlled for).

Beyond the complexity introduced by carry and borrow operations, research has revealed systematic differences between arithmetic operations like addition and subtraction. Educational studies showed that subtraction is particularly problematic for many students (Narciss & Huth, 2006; Riccomini, 2005). Specifically, solving subtraction problems is more prone to errors and takes longer than solving addition problems (Artemenko, Soltanlou, Dresler, et al., 2018; Yao et al., 2025). Neuroimaging studies have demonstrated that both addition and subtraction operations recruit the fronto-parietal network (Istomina & Arsalidou, 2024). However, some research showed that subtraction elicits more bilateral activation across frontal and parietal cortices, including the IPS and the supramarginal gyrus (SMG, e.g., Montefinese et al., 2017), suggesting higher cognitive demands (Rosenberg-Lee et al., 2011). These findings suggest more intensive recruitment of neural resources during subtraction. While subtraction may involve more diffuse or bilateral activation, findings that link addition to left-lateralized activity in the angular gyrus (AG) typically use single-digit addition problems that might be solved by fact-retrieval (Grabner et al., 2009), which should not be uncritically generalized to multi-digit arithmetic. These findings (Artemenko et al., 2015; Artemenko, Soltanlou, Dresler, et al., 2018; Yi-Rong et al., 2011) indicate heterogeneous lateralization across studies and underscore the need for further investigation of arithmetic operations, particularly within multi-digit arithmetic.

Arithmetic requires working memory – especially for the carry and borrow operations, which necessitate the temporary storage and manipulation of intermediate results (Imbo et al., 2005). Dual-task and interference studies using two-digit problems show that particularly carry-related performance declines under phonological and visuospatial working memory loads, indicating the involvement of multiple working memory subsystems in supporting place-value computations (Dai, 2018; Imbo et al., 2005; Nuerk et al., 2001). Besides, neural research showed that higher verbal working memory capacity was associated with a larger neural carry effect in the left IFG (i.e., greater differential activation for carry vs. non-carry problems) and improved behavioral outcomes, suggesting that individuals with higher working memory capacity flexibly recruit additional frontal resources when faced with increased arithmetic difficulty (Artemenko et al., 2018). This indicates that higher working memory capacity allows for the recruitment of frontal resources to meet increased arithmetic demands, but it is unclear how the underlying processes are for multi-digit arithmetic, which is even more complex.

Although previous behavioral and neuroimaging studies indicate that carry and borrow operations increase working memory demands and recruit frontal and parietal regions in two-digit arithmetic, it remains unclear whether these effects generalize to three-digit arithmetic where multiple carry/borrow steps can co-occur within a single problem. The current study aims to investigate whether the difficulty of multi-digit arithmetic is mainly processed in domain-general frontal areas or also in domain-specific parietal areas. Given the limitations of traditional neuroimaging techniques, the current study uses fNIRS to assess multi-digit arithmetic in an ecologically valid production paradigm (cf. Artemenko, 2018; Yao et al., 2025). Based on previous research, we preregistered (https://osf.io/ef3aw/) three hypotheses using the preregistration template for fNIRS studies (Schroeder et al., 2023): Regarding arithmetic complexity, we expect a carry effect in multi-digit addition and a borrow effect in multi-digit subtraction, with arithmetic performance decreasing and brain activation increasing (especially in left IFG, bilateral MFG, and left IPS) as the number of carry or borrow operations increases (0, 1, 2). Regarding arithmetic operation, we expect performance in subtraction to be worse than in addition, with subtraction eliciting more bilateral fronto-parietal activation, whereas addition is expected to be predominantly left-lateralized. Furthermore, we expect the carry and borrow effects in multi-digit arithmetic to be associated with working memory both on a behavioral (performance) and on a neural level (frontal activation), beyond the general relation between multi-digit arithmetic and working memory.

Methods

Participants. A total of $N = 55$ adults (19 male, 36 female; age: $M = 23.98$ years, $SD = 3.85$ years) were recruited via the university's list mailing system. All participants were right-handed, aged between 18 and 40 years, had normal or corrected-to-normal vision, and reported no history of neurological, mental or learning disorders. Following pre-specified exclusion criteria, participants were excluded due to drop out ($n = 1$), due to low task performance (accuracy < 50%; $n = 0$) or reaction time (RT) outliers (*Median* ± 3 *MAD*; $n = 6$). Furthermore, additional participants were excluded from the fNIRS data analysis due to insufficient data (> 50% missing trials per condition; $n = 0$) or poor signal quality (> 3 noisy channels; $n = 2$). Thus, the final sample consisted of $N = 48$ participants for behavioral analysis and $N = 46$ participants for neural analysis. For participation, participants received either course credits or monetary compensation. The study was approved by the local ethics committee for psychological research of the University of Tuebingen and was in accordance with the latest version of the Declaration of Helsinki.

Materials. The mental arithmetic task consisted of three-digit addition and subtraction problems, each categorized into three levels of complexity. For addition, three levels of complexity were defined according to the number of carry operations (0, 1, 2): no carry operation, 1 carry operation either at the decade position or at the hundred positions, carry operations at both the decade and hundred positions (see Table 1 for examples). For subtraction, three levels of complexity were defined according to the number of borrow operations (0, 1, 2), corresponding to inverse addition problems (also see Table 1 for examples). All conditions were matched for stimulus properties, i.e., the mean of the two addends was matched across conditions ($M = 375$ for each addend, $M = 750$ for the sum), and the position of the larger addend was counterbalanced. Each arithmetic operation included 96 trials, in equal parts including problems without carry/borrow, with carry/borrow at the decade position only, with carry/borrow at the hundred position only, and with carry/borrow at both the decade and the hundred positions. For analysis, the decade-only and hundred-only carry/borrow problems were merged into the complexity 1 level.

The mental arithmetic task was programmed using OpenSesame (version 3.3.10). Participants were instructed to solve each problem as accurately and quickly as possible in a computerized written production paradigm, by typing their answers into a number keyboard. Each trial began with a fixation dot presented for 500 ms, followed by the arithmetic problem, which remained on screen until the participant started a response. During the response, the problem disappeared, the entered digits for the response were neither visible nor editable, and there was no time limit for typing the response. Trials were separated by a 10 s inter-trial interval, jittered between 9 and 11 s.

Table 1. Arithmetic problem examples for addition and subtraction varying in complexity.

complexity	addition	subtraction
0 (no carry/borrow)	425 + 512	937 – 425
1 (carry/borrow at the decade position)	548 + 236	784 – 548
1 (carry/borrow at the hundred position)	246 + 482	728 – 246
2 (carry/borrow at both positions)	152 + 569	721 – 152

Procedures. Participants completed two experimental sessions on separate days, separating the different arithmetic operations (addition vs. subtraction). The order of sessions was counterbalanced across participants and the order of trials randomized within each session. During each session, participants performed the arithmetic task while brain activation was recorded using fNIRS in a dimly lit room. After the second session, participants completed both forward and backward versions of the Letter Span Test and the Corsi Block-Tapping Test (see also in Artemenko et al., 2018). The forward version of the Letter Span Test assesses verbal short-term memory and the backward version verbal working memory (Conway et al., 2005). Similarly, the forward version of the Corsi Block-Tapping Test assesses visuospatial short-term memory and the backward version visuospatial working memory (Berch et al., 1998). Additionally, state and trait math anxiety, test anxiety, math self-concept, and math motivation were assessed via standardized questionnaires, but not considered in the current study.

fNIRS data acquisition. fNIRS measures cortical activation by means of changes in oxygenated (HbO) and deoxygenated haemoglobin (HbR) concentrations (Ferrari & Quaresima, 2012; Scholkmann et al., 2014). Brain activation was recorded using the ETG-4000 Optical Topography System (Hitachi Medical Corporation, Tokyo, Japan) in a dimly lit room. Continuous wave laser diodes with wavelengths of 695 ± 20 nm and 830 ± 20 nm were used as light sources, with a sampling rate of 10 Hz. The optode arrangement comprised 10 sources and 8 detectors, forming 16 channels covering frontal and parietal brain areas in both hemispheres: 3 parietal channels (IPS, SMG, AG) and 5 frontal channels (MFG, IFG) per hemisphere (see Figure 1). The channels were mapped on the brain based on standard virtual registration methods (Tzourio-Mazoyer et al., 2002) using the automated anatomic labeling (AAL) atlas.

Data analysis.

Statistical analysis

Behavioral data analysis. Behavioral data were analysed using 2 × 3 repeated-measures ANOVAs with operation (addition vs. subtraction) and complexity (0, 1, 2 carry/borrow operations) as within-subject factors. Accuracy (ACC) and reaction time (RT) were used as dependent measures. The analyses were conducted using JASP (Jeffrey's Amazing Statistics Program, version 0.95, JASP Team, 2025) (https://jasp-stats.org/). Trials were excluded in case of unexpected events (0% for both addition and subtraction), RTs shorter than 200 ms (anticipatory responses, 0%), or RTs that deviated more than ±3 median absolute deviations (*MAD*) from the subject's *median* RT (2.93% for addition and 2.45% for subtraction). Accuracy was calculated as the proportion of correct responses among the remaining valid trials.

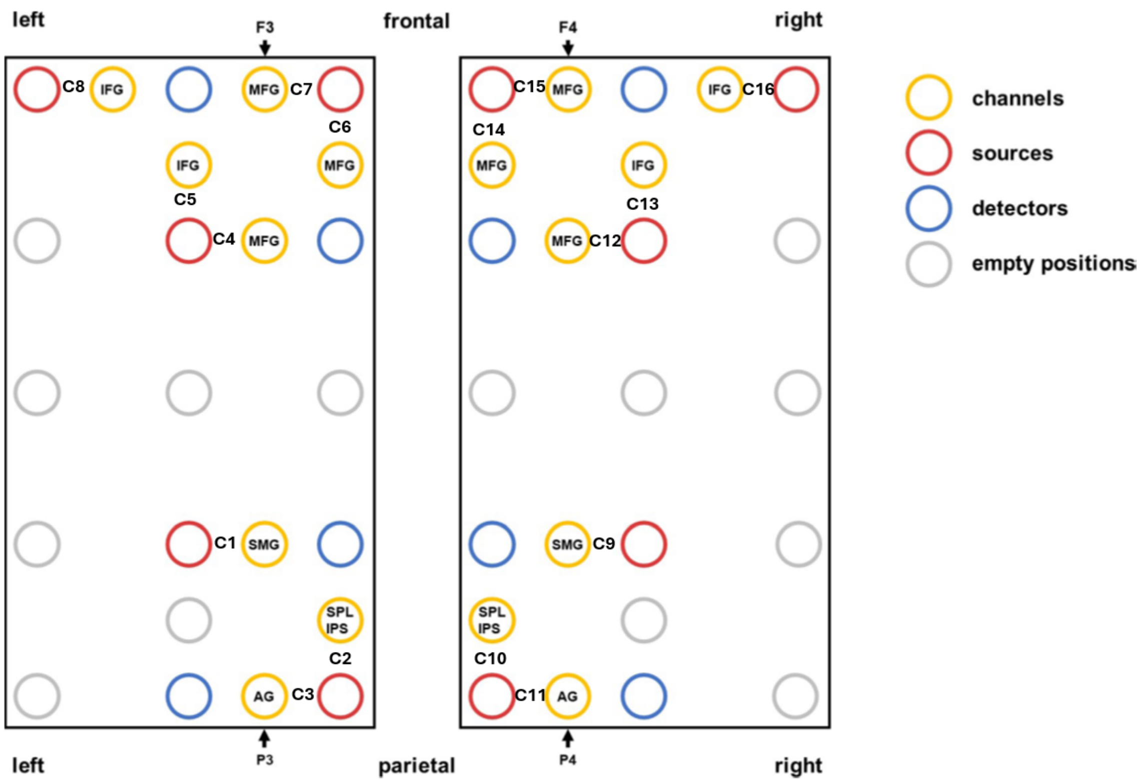
fNIRS analysis. fNIRS data were analysed by using the NIRS Brain AnalyzIR Toolbox (Santosa et al., 2018) based on MATLAB (The Mathworks Inc, Natick, MA, USA). As preregistered, the fNIRS data were preprocessed using a standard pipeline, including visual inspection, temporal derivative distribution repair (TDDR) for motion artifact correction, and low pass filtering at 0.2 Hz. Optical density data were converted into concentration changes in oxygenated (ΔHbO) and deoxygenated (ΔHbR) haemoglobin via the modified Beer–Lambert law (Scholkmann et al., 2013). To further account for interindividual variability, each subject's haemoglobin time series was z-scored on a channel-wise basis. Although this step was not preregistered, we considered it essential for ensuring comparability across participants.

Task-related responses were estimated at the subject level using a canonical hemodynamic response function within an AR-IRLS GLM framework with prewhitening and third-order Legendre polynomial drift removal (AR-IRLS; Barker et al., 2016). The canonical peak-time parameter was set to 9 s, based on the group average peak latency, to achieve optimal model fit. Resulting β-weights for ΔHbO and ΔHbR were exported for each subject, channel and condition. At the group level, we performed 2 × 3 repeated-measures ANOVAs with operation (addition vs. subtraction) and complexity (0, 1, 2) on channel-wise β estimates for ΔHbO and ΔHbR using custom MATLAB scripts, with FDR-correction across all channels ($\alpha = .05$). No significant effects were observed in this preregistered confirmatory analysis (see Table S3 & S4 in Supplementary Materials). One possible explanation is that model-based approaches such as the GLM, which rely on a predefined hemodynamic response function, may be less sensitive to condition-related differences when interindividual variability in response timing or shape is high. Another possibility is that β-values were uniformly high across all conditions, such that no differences could be detected in this cognitively demanding task.

In fNIRS research, block average is widely regarded as a standard and robust method (Ortega-Martinez et al., 2022), as it avoids assumptions about the hemodynamic response function and is less affected by parameter choices compared to GLM or FIR models. In line with this interpretation but deviating from the preregistration, we changed our analytic approach and chose to analyse the mean amplitude in a block-average approach, which is meant to be more robust to detect activation changes (Ortega-Martinez et al., 2022). In a block-average approach, haemoglobin responses were analysed in the time window [−3 20] s, with mean amplitudes determined for HbO and HbR within [5 15] s after stimulus onset, baseline-corrected to the 3 s preceding stimulus onset. The same 2 × 3 ANOVAs with operation (addition vs. subtraction) and complexity (0, 1, 2) were conducted for mean amplitude for each channel. FDR-correction was applied across all channels ($\alpha = .05$), yielding q values. Post-hoc pairwise comparisons for the main effect of complexity were conducted, their p values were Bonferroni-adjusted within channel across the three contrasts.

Brain-behaviour correlations. To assess brain-behaviour relationships, Pearson correlation analyses were conducted between individual working memory capacity (verbal and visuospatial short-term and working memory) and behavioral and neural complexity effects using JASP. Behavioral complexity effects reflect differences in reaction time and accuracy across levels, with the difference from complexity level 0 to 1 capturing the first carry/borrow effect and the shift from level 1 to 2 capturing the second carry/borrow effect. The carry/borrow effects were calculated for ACC ($\Delta ACC_{0\text{ to }1} = ACC_0 - ACC_1$ and $\Delta ACC_{1\text{ to }2} = ACC_1 - ACC_2$) and RT ($\Delta RT_{0\text{ to }1} = RT_1 - RT_0$ and $\Delta RT_{1\text{ to }2} = RT_2 - RT_1$). Neural complexity effects reflect mean changes in HbO and HbR. For HbO, the carry/borrow effects were calculated as $\Delta HbO_{0\text{ to }1} = HbO_1 - HbO_0$ and $\Delta HbO_{1\text{ to }2} = HbO_2 - HbO_1$. For HbR, the effects were calculated in the opposite direction, i.e., $\Delta HbR_{0\text{ to }1} = HbR_0 - HbR_1$ and $\Delta HbR_{1\text{ to }2} = HbR_1 - HbR_2$. Deviated from our preregistration, we initially planned to perform multiple regression analyses including complexity levels (0,1,2) and four memory measures (verbal/visuospatial short-term/working memory) as predictors, together with their interactions, for all dependent variables. In the final analyses, we instead computed subject-level complexity deltas (ΔACC/ ΔRT/ ΔHbO/ ΔHbR) and correlated them with memory scores. This change was made to avoid collinearity among the memory measures and instability of higher-order interactions given the sample size, and to better capture our theoretical focus on between-person modulation of within-person complexity effects. Inference thus concerns associations with summary complexity effects rather than trial-level interactions.

Figure 1. fNIRS probe sets covering the bilateral fronto-parietal brain network



Notes. The figure is adapted from Artemenko (2021). The probesets were fixed at P3/P4 and oriented towards F3/F4; the positions of channels and optodes (sources and detectors) including empty positions are marked. *Abbreviations of the channel labels*: IFG – inferior frontal gyrus, MFG – middle frontal gyrus, SPL/IPS – superior parietal lobule/intraparietal sulcus, SMG – supramarginal gyrus, AG – angular gyrus.

Results

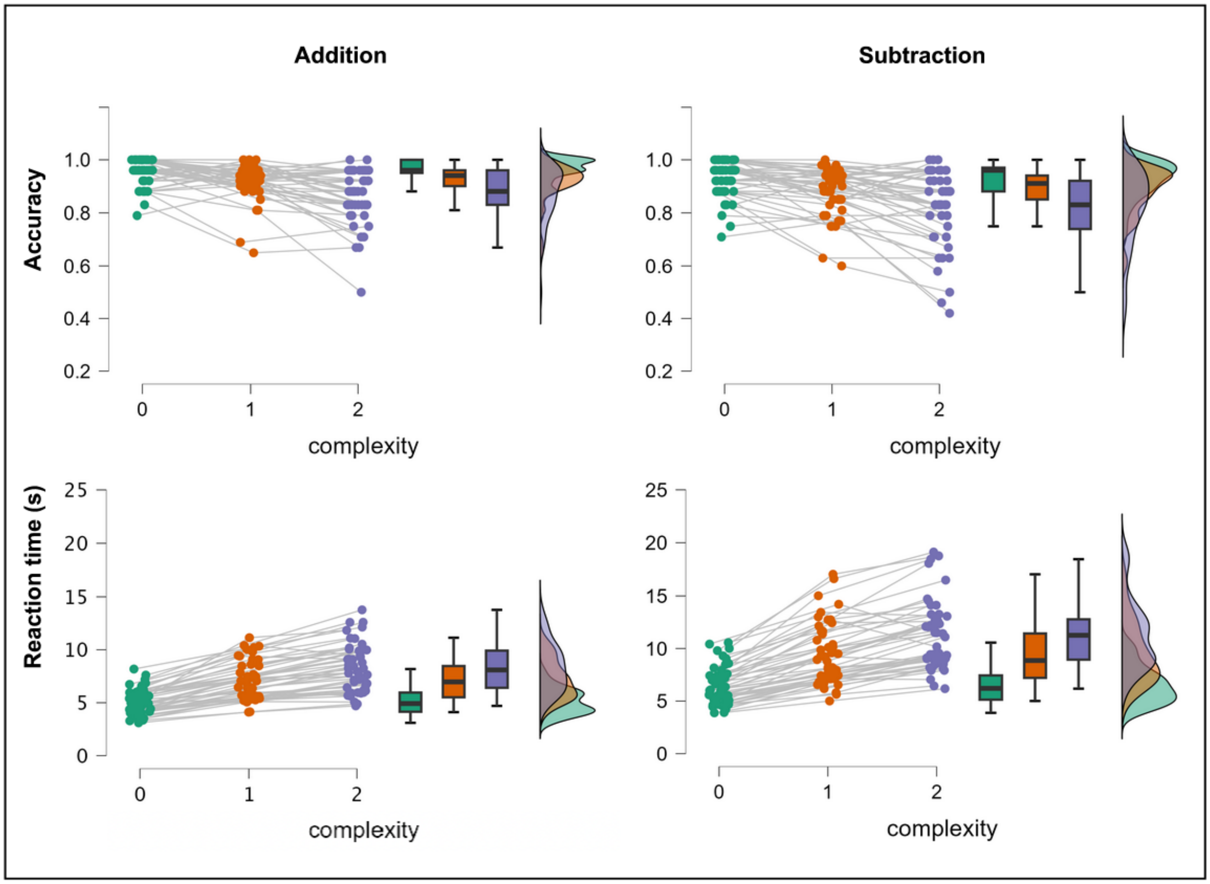
Behavioral results

The descriptive data for accuracy and reaction time can be found in Supplementary Materials, Table S1. Data, analysis scripts, and materials will be openly shared at https://osf.io/ef3aw/overview following publication.

The ANOVA on accuracy revealed significant main effects of operation, $F(1, 47) = 17.79$, $p < .001$, $\eta^2_p = 0.275$, and complexity, $F(2, 94) = 38.95$, $p < .001$, $\eta^2_p = 0.453$ (see Figure 2). Accuracy was generally higher ($t = 4.22$, $p < .001$, *Cohen's d* $= 0.42$) for addition ($M = 0.92$, $SD = 0.08$) compared to subtraction ($M = 0.88$, $SD = 0.11$). Additionally, accuracy declined as complexity increased from 0 ($M \pm SD = 0.96 \pm 0.05$) to 1 ($M \pm SD = 0.92 \pm 0.07$; $p < .001$) ($t = 5.4$, $p < .001$, *Cohen's d* $= 0.45$) and from 1 to 2 ($M \pm SD = 0.87 \pm 0.10$; $p < .001$) ($t = 4.74$, $p < .001$, *Cohen's d* $= 0.61$) carry/borrow operations. The interaction between operation and complexity was not significant, $F(2, 94) = 1.36$, $p = .262$, $\eta^2_p = 0.028$.

The ANOVA on reaction times revealed significant main effects of operation, $F(1, 47) = 57.33$, $p < .001$, $\eta^2_p = 0.550$, and complexity, $F(2, 94) = 241.44$, $p < .001$, $\eta^2_p = 0.837$. Reaction times were faster ($t = 7.57$, $p < .001$, *Cohen's d* = 0.96) for addition ($M \pm SD$ = 6.84 s ± 1.81) than for subtraction ($M \pm SD$ = 9.06 s ± 2.71). Additionally, reaction time increased as complexity increased from 0 ($M \pm SD$ = 5.11 s ± 1.20) to 1 ($M \pm SD$ = 7.09 s ± 1.82; $p < .001$) ($t = 14.97$, $p < .001$, *Cohen's d* = 1.07) and from 1 to 2 ($M \pm SD$ = 8.33 s ± 2.25; $p < .001$) ($t = 11.35$, $p < .001$, *Cohen's d* = 0.67) carry/borrow operations. Furthermore, the interaction between operation and complexity was significant, $F(2, 94) = 22.90$, $p < .001$, $\eta^2_p = 0.327$, indicating that the increase in reaction times across complexity levels differed between addition and subtraction. The first carry effect from complexity level 0 to 1 was significantly larger than the second carry effect from complexity level 1 to 2 in addition ($M \pm SD_{\text{carry effect 0 to 1}}$ = 1.99 s ± 0.96; $M \pm SD_{\text{carry effect 1 to 2}}$ = 1.24 s ± 0.86; $t = 5.16$, $p < .001$, *Cohen's d* = 0.74). Likewise, the first borrow effect from complexity level 0 to 1 was significantly larger than the second borrow effect from complexity level 1 to 2 ($M \pm SD_{\text{borrow effect 0 to 1}}$ = 2.93 s ± 1.57; $M \pm SD_{\text{borrow effect 1 to 2}}$ = 1.83 s ± 1.51; $t = 3.46$, $p = .001$, *Cohen's d* = 0.50). Critically, borrow effects were significantly larger than the corresponding carry effects at both steps (0 to 1: $t = 5.13$, $p < .001$, *Cohen's d* = 0.74; 1 to 2: $t = 2.59$, $p = .013$, *Cohen's d* = 0.37), explaining the observed interaction between operation and complexity.

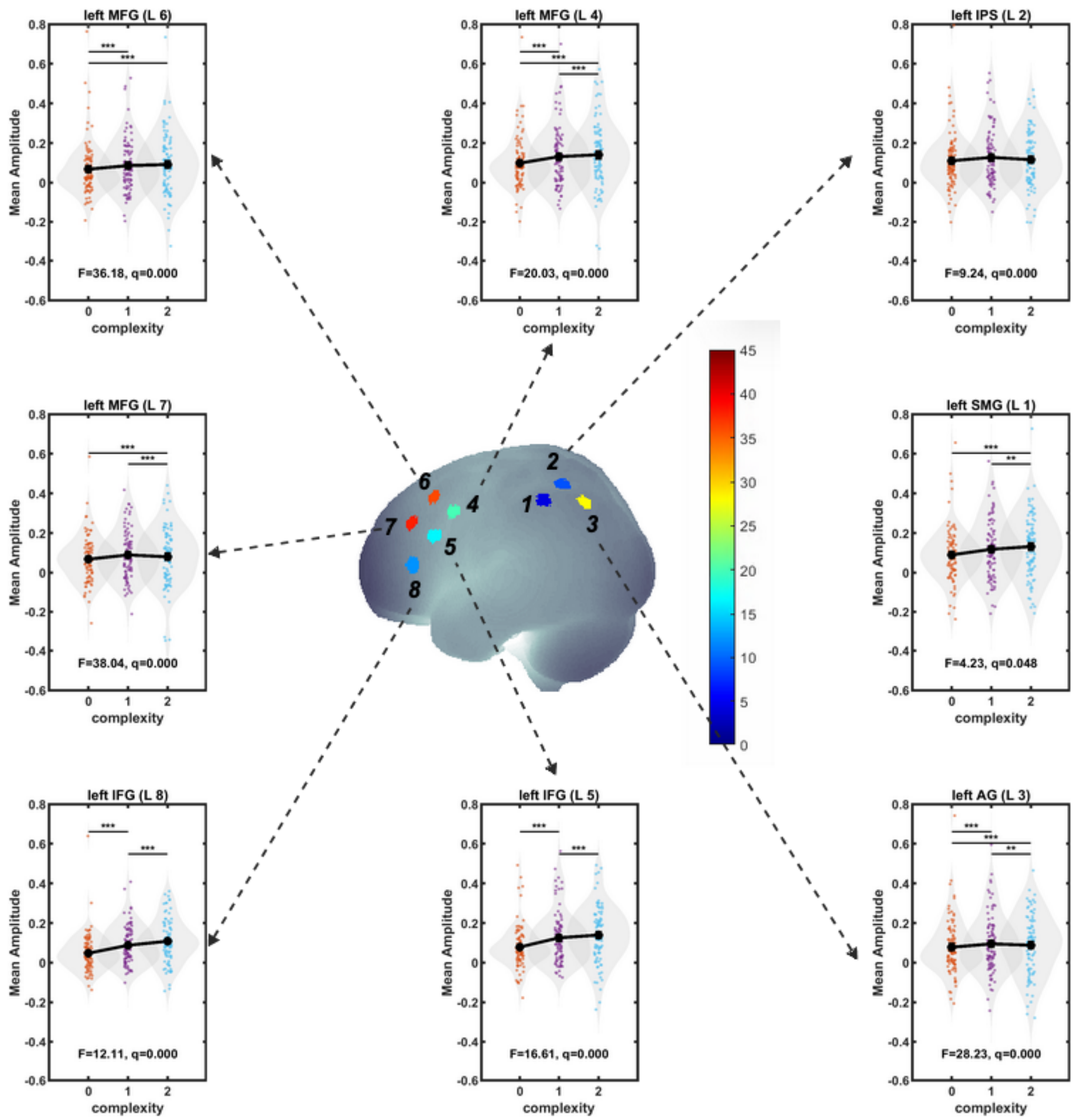
Figure 2. Behavioral performance in addition and subtraction depending on complexity



fNIRS results

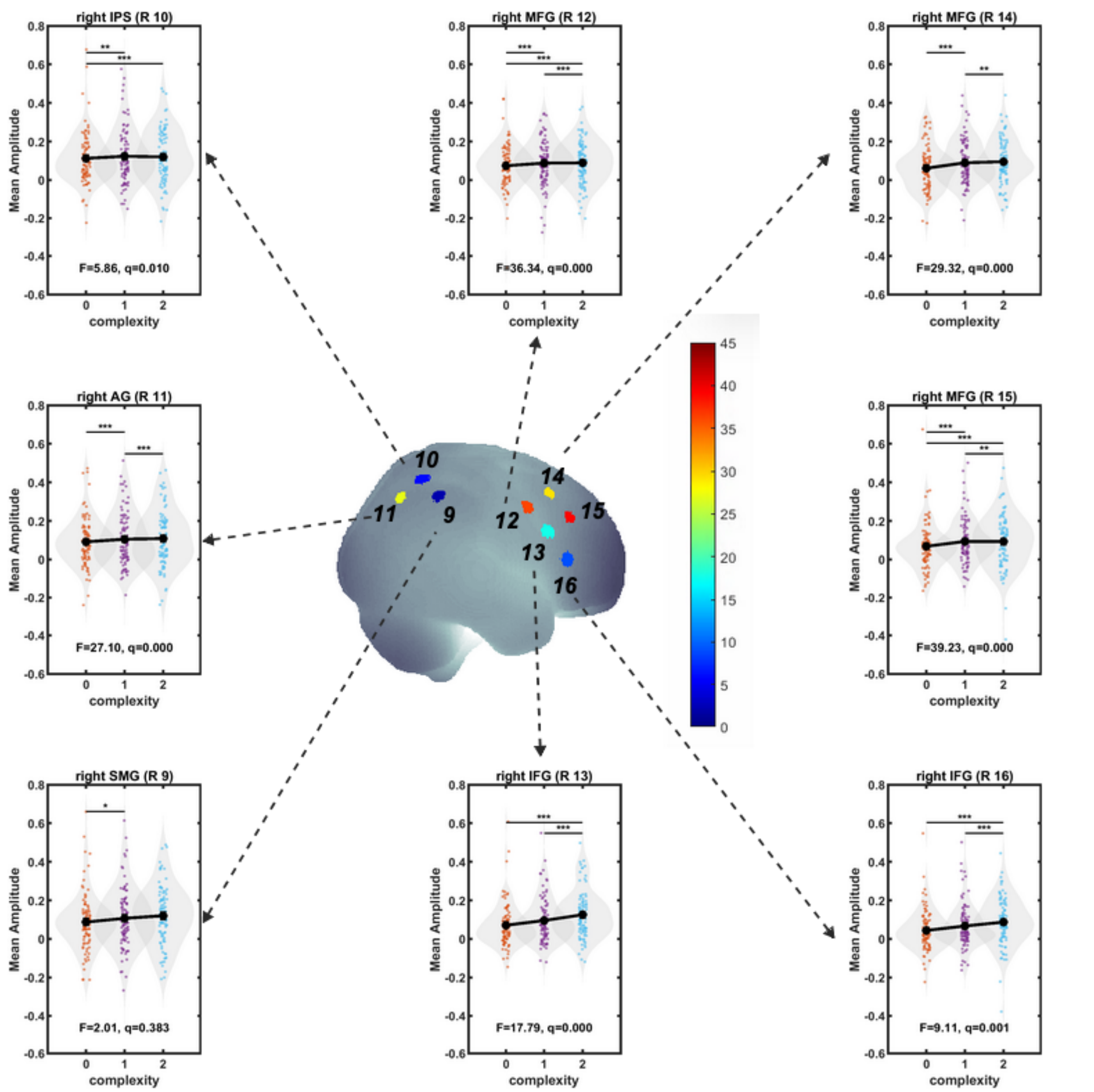
For mean amplitude, significant main effects of complexity were observed in frontal areas (bilateral IFG and MFG) and parietal areas (bilateral IPS, left SMG and bilateral AG) for both HbO and HbR, and additionally in the right SMG for HbR only (all $q < .05$, see Table S5 & S6 in Supplementary Materials). No significant main effects of operation (addition vs. subtraction) or interaction effects between operation and complexity were observed in any channel for either HbO or HbR (all $q > .05$, see Table S5 & S6 in Supplementary Materials). Post-hoc comparisons are shown in Figure 3 & 4 for HbO and Figure S1 & S2 for HbR. Overall, in the frontal cortex, bilateral MFG and left IFG showed increasing trend for mean amplitude of HbO and decreasing trend for mean amplitude of HbR with complexity level increasing from 0 to 2. Right IFG only showed mean amplitude of HbO increase and HbR decrease at level 2, whereas level 0 to 1 was not significantly different. In the parietal cortex, there was no consistent trend across regions. Specifically, no pairwise contrast survived FDR correction in left IPS, whereas in the right IPS, HbO increased and HbR decreased from level 0 to 1, with no additional significant change from 1 to 2. Right AG showed increasing trend for mean amplitude of HbO and decreasing trend for mean amplitude of HbR with complexity level increasing from 0 to 2.

Figure 3. Fronto-parietal activation for complexity effects in the left hemisphere (HbO mean amplitude)



Notes. Colors on the cortical map denote the F statistic of the main effect of complexity (0/1/2) from two-way ANOVAs on HbO mean amplitude at each left-hemisphere channel (warmer colors = larger F). Violin plots show participant distributions; black dots indicate means connected by lines; asterisks mark Bonferroni-adjusted post-hoc pairwise contrasts (*** $p < .001$, ** $p < .01$, * $p < .05$). Descriptive means ± SD for all channels and levels are provided in Table S2.

Figure 4. Fronto-parietal activation for complexity effects in the right hemisphere (HbO mean amplitude)



Notes. Colors on the cortical map denote the F statistic of the main effect of complexity (0/1/2) from two-way ANOVAs on HbO mean amplitude at each right-hemisphere channel (warmer colors = larger F). Violin plots show participant distributions; black dots indicate means connected by lines; asterisks mark Bonferroni-adjusted post-hoc pairwise contrasts (*** $p < .001$, ** $p < .01$, * $p < .05$). Descriptive means ± SD for all channels and levels are provided in Table S2.

Brain-behavior correlations

Significant correlations were found between the complexity effects and working memory. At the behavioral level, higher verbal short-term memory and working memory were significantly correlated with a smaller complexity effect in terms of $\Delta ACC_{1\ \mathrm{to}\ 2}$ (all $p < .05$, see Table S7 in Supplementary Materials). Additionally, higher visuospatial working memory was significantly correlated with smaller complexity effect in terms of $\Delta ACC_{0\ \mathrm{to}\ 1}$. At the neural level, for HbO, higher verbal short-term memory was related to reduced complexity effects in bilateral MFG ($\Delta HbO_{1\ \mathrm{to}\ 2}$). Higher visuospatial short-term memory was related to reduced complexity effects in the right SMG and AG ($\Delta HbO_{1\ \mathrm{to}\ 2}$; all $p < .05$, see Table S8 in Supplementary Materials). For HbR, higher verbal short-term memory was related to increased complexity effects in left SMG ($\Delta HbR_{0\ \mathrm{to}\ 1}$) and higher verbal working memory was related to reduced complexity effects in the bilateral SMG ($\Delta HbR_{1\ \mathrm{to}\ 2}$). Higher visuospatial short-term memory was related to reduced complexity effects in right MFG ($\Delta HbR_{0\ \mathrm{to}\ 1}$) and SMG ($\Delta HbR_{1\ \mathrm{to}\ 2}$); higher visuospatial working memory was related to increased ($\Delta HbR_{0\ \mathrm{to}\ 1}$) or reduced ($\Delta HbR_{1\ \mathrm{to}\ 2}$) complexity effects in left IFG (all $p < .05$, see Table S9 in Supplementary Materials). All other correlations were not significant (see Table S8 & S9 in Supplementary Materials). Collectively, these findings indicate that individual differences in working memory capacity modulate fronto-parietal hemodynamic responses to arithmetic complexity, with the modulation depending on both the specific brain region involved and the type of working memory engaged.

Discussion

In this fNIRS study on multi-digit arithmetic, we investigated whether increases in arithmetic complexity (number of carry/borrow operations) predominantly rely on domain-general frontal resources or additionally recruit domain-specific parietal magnitude/place-value systems. We found that increasing complexity impaired performance (lower accuracy and longer reaction time), with stronger reaction time costs for subtraction. At the neural level, higher complexity engaged the fronto-parietal network more strongly, particularly in the bilateral IFG and MFG as well as the right IPS and AG. Working memory differentially buffered these complexity effects: individuals with higher verbal short-term/working memory were better in more complex problems and needed less frontal resources (MFG), whereas individuals with higher visuospatial memory needed less parietal resources (SMG/AG).

Processing demands of arithmetic complexity

Our findings indicate strong effects of arithmetic complexity on behavioral performance and the associated brain activation during multi-digit arithmetic. With increasing number of carry and borrow operations from 0 to 2, each additional step required extra computation in three-digit arithmetic, leading to longer reaction times and a higher likelihood of errors (Deschuyteneer et al., 2005; Imbo et al., 2007b). This increased difficulty elicited greater activation in the fronto-parietal network of arithmetic processing, including frontal activation in bilateral MFG and IFG as well as parietal activation in SMG, AG and right IPS.

The widespread bilateral activation observed in our study may attribute to the high difficulty and processing demands of multi-digit arithmetic, which engage both domain-general frontal and domain-specific parietal resources. According to the two-network model of numerical cognition (Klein & Knops, 2023), the frontal cortex, particularly the IFG, supports domain-general functions such as strategy selection, executive control, and the coordination of multi-step procedures (Owen et al., 2005). The MFG is implicated in computations requiring procedural steps like carrying, correlating with working memory demands and increased cognitive control (Arsalidou & Taylor, 2011; Verner et al., 2013). These findings align with our observation that HbO increased (and HbR decreased) in bilateral MFG and IFG with task difficulty, reflecting heightened coordination, working memory and control efforts.

In contrast, parietal regions are functionally specialized: the IPS is crucial for number magnitude processing and specifically right IPS is involved in place-value processing (Artemenko et al., 2015). The AG and SMG are key components of the network for arithmetic fact retrieval (Klein & Knops, 2023). Prior neuroimaging work in two-digit arithmetic has shown that carry/borrow manipulations modulate IPS activity. While a predominantly left-hemisphere involvement in the IPS is observed in several studies for such operations (e.g., Artemenko et al., 2018; Klein et al., 2010; Kong et al., 2005), it is important to note that lateralization can vary with task demands and problem complexity. For instance, the right IPS becomes more strongly engaged under the most challenging combinations of problem size and carry operations (Klein et al., 2010). Our results extend this literature to multi-digit arithmetic and indicate stronger recruitment of the right IPS, consistent with additional place-value integration demands in more complex computations.

Importantly, problem size was counterbalanced in our design, ruling out numerical magnitude as an explanation, which would recruit the left IPS (Ansari, 2007; Bahreini et al., 2023), and indicating that the observed effects stem from the additional place-value integration required by carry and borrow operations, primarily involving the right IPS (Artemenko et al., 2015). This interpretation is supported by our neural pattern, with a selective right-IPS HbO increase and HbR decrease when complexity increased from level 0 to level 1. Notably, the right IPS activation in our data is categorical: activation increased from no carry/borrow (level 0) to carry/borrow (level 1), but did not rise further from level 1 to 2. This suggests that the right IPS indicates the need for place-value integration rather than the number of carry/borrow steps, consistent with previous findings that distinguish categorical from continuous aspects of the carry effect and with causal evidence implicating the right parietal cortex in place-value processing (Klein et al., 2010). Additional carry/borrow steps may instead load domain-general control and working memory supported by prefrontal cortex or the right AG, which could explain the absence of further right IPS increases.

Although the fact network is typically engaged in single-digit arithmetic, multi-digit operations also rely on arithmetic facts for intermediate steps (sum/difference of the units, decades, hundreds), explaining its involvement. Taken together, the processing demands associated with increasing arithmetic complexity in multi-digit arithmetic consist of procedural problem-solving and working memory processes in the IFG and MFG, place-value integration processes in the right IPS, and arithmetic fact knowledge in the AG/SMG.

In summary, higher arithmetic complexity in multi-digit arithmetic necessitates coordinated engagement of domain-general and domain-specific processes in the fronto-parietal network, and these effects are further modulated by individual working-memory capacity.

The role of working memory

Beyond the general complexity effects, our findings revealed that individual working memory capacity plays a selective buffering role. Behaviorally, higher verbal short-term and working memory capacity were associated with smaller accuracy losses from complexity level 1 to 2, suggesting that verbal working memory resources (e.g., the phonological loop) facilitate the retention and manipulation of intermediate results (e.g., the "carry-over" digit). Higher visuospatial working memory were associated with smaller accuracy losses from zero to one carry/borrow steps, indicating support at the onset of carry/borrow when column alignment and place-value mapping are first required (see Table S7). These associations are consistent with prior work linking working memory to performance in complex arithmetic and carry/borrow demands (Artemenko, et al., 2018; DeStefano & LeFevre, 2004).

Neurally, higher verbal short-term memory was linked to weaker complexity-related increases in bilateral MFG activation. This pattern is consistent with the idea that individuals with higher verbal maintenance capacity require less additional frontal recruitment under increasing carry/borrow demands, in line with evidence that frontal cortex (including IFG/MFG) support multi-step coordination and executive control during difficult arithmetic (Artemenko, et al., 2018).

In contrast, higher visuospatial short-term memory was associated with weaker increases in parietal regions (right SMG and AG;). This suggests reduced additional parietal recruitment in individuals with stronger visuospatial resources, consistent with visuospatial involvement in multi-digit formatting/alignment and with parietal spatial circuitry engaged during mental arithmetic;

note that we report visuospatial and verbal associations separately without implying direct differences between them (Trbovich & LeFevre, 2003).

Taken together, these dissociable associations suggest complementary contributions of working-memory components to complex multi-digit arithmetic: verbal mechanisms may help buffer and update carry/borrow digits across steps, whereas visuospatial resources may mitigate the initial cost of introducing a carry/borrow (formatting and place-value mapping). This functional segregation aligns with established knowledge about working memory and with behavioral evidence on distinct verbal and visuospatial roles in complex calculation (Baddeley, 1992; Imbo & LeFevre, 2010). It also accords with findings that working-memory capacity predicts mathematical competence (Bull et al., 2008; Zhang et al., 2019). While another dual-task study have emphasized a predominant visuospatial contribution in multi-digit arithmetic (Clearman et al., 2017), our results point to complementary roles of both verbal and visuospatial short-term memory as problem complexity increases. These interpretations are correlational and do not imply causality.

Similarities and differences between addition and subtraction

Replicating previous findings (Artemenko et al., 2018; Campbell et al., 2006; Klein et al., 2014), both addition and subtraction showed clear arithmetic complexity (carry/borrow) effects. Moreover, borrow effects were larger than carry effects. Solving subtraction problems takes longer and is more prone to errors than solving addition problems (Campbell, 2008), a disparity that was further amplified as problem complexity increased.

Given the typically higher cognitive demands of subtraction compared to addition, the observed behavioral pattern might intuitively suggest that subtraction would induce stronger neural activation in the brain regions for multi-digit arithmetic processing. However, our fNIRS results did not reveal significant operation-specific activation differences within the fronto-parietal network. Note that problem size was controlled and our paradigm utilized multi-digit tasks, inherently mandating algorithmic, stepwise computation over fact retrieval, thereby likely minimizing the strategy variations observed in simpler problems (De Smedt et al., 2011). These controls indicate that, for three-digit arithmetic of comparable complexity, addition and subtraction engage similar fronto-parietal resources in terms of cortical activation intensity (cf. Arsalidou & Taylor, 2011; Dehaene et al., 2003).

Our results of no significant operation-specific activation differences in the measured fronto-parietal network between addition and subtraction do not rule out operation-specific contributions from regions beyond our fNIRS coverage. Previous neuroimaging studies have shown that addition and subtraction can engage partly distinct neural networks. For instance, Kong et al. (2005) reported separable cortical networks for the two operations, with subtraction recruiting additional cortical regions, but no interaction between operation type and carry/borrow complexity. Similarly, Yi-Rong et al. (2011) demonstrated partially distinct addition and subtraction networks, with subtraction involving subcortical structures, and purposely designed their study to avoid confounding by carry/borrow factors. Other work has implicated anterior cingulate cortex and the medial prefrontal cortex in arithmetic operations (Istomina & Arsalidou, 2024), basal ganglia in rule-based/procedural aspects of calculation (Gullick & Wolford, 2012), and the cerebellum during addition/subtraction and symbolic arithmetic (Ashburn et al., 2024; Saban et al., 2024). Moreover, meta-analytic work supports the idea that different arithmetic operations may rely on distinct sets of brain regions or network interactions beyond the fronto-parietal network when different strategies are employed (Istomina & Arsalidou, 2024; Yang et al., 2017). For instance, simple addition often relies on fact retrieval, engaging left angular gyrus (Grabner & De Smedt, 2011; Sokolowski et al., 2023), whereas subtraction and complex operations lean more towards procedural steps, recruiting areas like IPS, premotor cortex, and other regions involved in working memory and sequencing (Campbell, 2008; Campbell & Xue, 2001; Yang et al., 2017). In our task, however, strategy variability was minimized by design; consequently, the stronger behavioral costs for subtraction most likely reflect the heavier procedural burden of borrow within this shared fronto-parietal network rather than distinct activation patterns. These findings underscore the need for future studies with broader neural coverage, including both cortical and subcortical regions.

Limitations and perspectives

The study's findings should be interpreted considering certain limitations. First, although fNIRS is well suited to measure cortical brain activation in ecological production paradigms, its limited spatial and depth resolution restricts measurement to the superficial cortex. Future work should also examine operation-related activity in arithmetic-relevant cortical and subcortical systems, such as the basal ganglia and cerebellum (Saban et al., 2024). Second, we did not track participants' strategies, and the computerized production paradigm might even have affected strategy use (e.g., if individuals prefer column-wise calculation from right-to-left but need to enter the response from left-to-right). However, production paradigms in mental arithmetic constrain strategy choice to calculation strategies rather than plausibility-checking or recognition-based responses as in decision paradigms (Yao et al., 2025). Future studies should assess or manipulate strategy use. Suitable approaches include eye-tracking in carry/borrow problems (Moeller et al., 2011) and trial-by-trial strategy reports or instructed strategy conditions (Polspoel et al., 2017), allowing tests of how strategies modulate brain activation and working-memory reliance. Finally, from a lifespan perspective, generalizability of the current findings remains to be tested: children and older adults may recruit partly different strategies and neural resources (e.g., hippocampal involvement during learning), so future work should include developmental and aging populations (Artemenko, 2021; Peters & De Smedt, 2017). As a perspective, computational modeling results correspond to our findings. Iglesias et al. (2025) showed that only networks trained on stepwise, decomposed curricula reproduced human-like carry effects. This supports a hierarchical learning process and suggests that efficient processing of multi-digit arithmetic, including carry and borrow, benefits from structured training in both human and artificial systems. Integrating such models with neuroimaging can inform adaptive instruction and clarify how symbolic arithmetic develops across the lifespan.

Conclusions

In conclusion, this fNIRS study provides neural evidence on how the brain manages the demands of complex three-digit arithmetic. Behaviorally, each additional carry or borrow step slowed responses and increased errors, with larger costs for subtraction. At the neural level, our findings confirm that multi-digit arithmetic complexity relies on domain-general working memory processes (IFG and MFG) as well as domain-specific place-value processes (IPS). Taken together, these results bridge the gap between two- and multi-digit arithmetic, refine the role of the right IPS as a categorical marker of place-value integration, and highlight that individual working-memory capacity contributes to neural efficiency under increasing arithmetic demands.

Author contributions: CRediT

Conceptualization: Xinru Yao, Christina Artemenko
Methodology: Xinru Yao, Beatrix Barth, Christina Artemenko
Software: Xinru Yao
Validation: Beatrix Barth
Formal analysis: Xinru Yao, Beatrix Barth, Christina Artemenko
Investigation: Xinru Yao
Resources: Xinru Yao
Data Curation: Xinru Yao
Writing – original draft: Xinru Yao
Writing – review & editing: Xinru Yao, Beatrix Barth, Christina Artemenko
Visualization: Xinru Yao
Supervision: Christina Artemenko
Project Administration: Xinru Yao
Funding Acquisition: Xinru Yao, Beatrix Barth, Christina Artemenko

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network (GSC1028, funded by the Excellence Initiative of the German federal and state governments). CA was supported by the German Research Foundation (DFG, grant number: 468460838, AR 1500/1-1; grant number: 513458453, AR 1500/2-1), European Social Fund and the Ministry of Science, Research and the Arts Baden-Wuerttemberg, and by the Tuebingen Postdoc Academy for Research on Education (PACE) at the Hector Research Institute of Education Sciences and Psychology. BB was supported by German Center for Mental Health (DZPG). We want to thank Annalena Wels for the data collection.

Data availability statement

Data, analysis scripts, and materials will be openly shared at https://osf.io/ef3aw/overview following publication.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the authors used OpenAI ChatGPT to improve the clarity and precision of the English language. The authors also used Moonlight (an AI-assisted literature-reading tool) solely to verify that cited references accurately matched the statements in the text. After using these tools, the authors reviewed and edited all content as needed and take full responsibility for the content of the published article.

References

Ansari, D. (2007). Does the Parietal Cortex Distinguish between “10,” “Ten,” and Ten Dots? *Neuron*, *53*(2), 165–167. https://doi.org/10.1016/j.neuron.2007.01.001

Arsalidou, M., & Taylor, M. J. (2011). Is 2 + 2 = 4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

Artemenko, C. (2018). *Neurocognitive Foundations of Arithmetic Complexity in Adults and Children*. Eberhard Karls Universität Tübingen.

Artemenko, C. (2021). Developmental fronto-parietal shift of brain activation during mental arithmetic across the lifespan: A registered report protocol. *PLOS ONE*, *16*(8), e0256232. https://doi.org/10.1371/journal.pone.0256232

Artemenko, C., Moeller, K., Huber, S., & Klein, E. (2015). Differential influences of unilateral tDCS over the intraparietal cortex on numerical cognition. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00110

Artemenko, C., Soltanlou, M., Dresler, T., Ehlis, A.-C., & Nuerk, H.-C. (2018). The neural correlates of arithmetic difficulty depend on mathematical ability: Evidence from combined fNIRS and ERP. *Brain Structure and Function*, *223*(6), 2561–2574. https://doi.org/10.1007/s00429-018-1618-0

Artemenko, C., Soltanlou, M., Ehlis, A.-C., Nuerk, H.-C., & Dresler, T. (2018). The neural correlates of mental arithmetic in adolescents: A longitudinal fNIRS study. *Behavioral and Brain Functions*, *14*(1), 1–13.

Ashburn, S. M., Matejko, A. A., & Eden, G. F. (2024). Activation and functional connectivity of cerebellum during reading and during arithmetic in children with combined reading and math disabilities. *Frontiers in Neuroscience*, *18*, 1135166. https://doi.org/10.3389/fnins.2024.1135166

Baddeley, A. (1992). Working Memory: The Interface between Memory and Cognition. *Journal of Cognitive Neuroscience*, *4*(3), 281–288. https://doi.org/10.1162/jocn.1992.4.3.281

Bahnmueller, J., Moeller, K., Mann, A., & Nuerk, H.-C. (2015). On the limits of language influences on numerical cognition – no inversion effects in three-digit number magnitude processing in adults. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01216

Bahreini, N., Artemenko, C., Plewnia, C., & Nuerk, H.-C. (2023). tDCS effects in basic symbolic number magnitude processing are not significantly lateralized. *Scientific Reports*, *13*(1), 21515. https://doi.org/10.1038/s41598-023-48189-z

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi Block-Tapping Task: Methodological and Theoretical Considerations. *Brain and Cognition*, *38*(3), 317–338. https://doi.org/10.1006/brcg.1998.1039

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*(3), 205–228. https://doi.org/10.1080/87565640801982312

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1005–1029. https://doi.org/10.1080/713756007

Campbell, J. I. D. (2008). Subtraction by addition. *Memory & Cognition*, *36*(6), 1094–1102. https://doi.org/10.3758/MC.36.6.1094

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, *130*(2), 299–315. https://doi.org/10.1037/0096-3445.130.2.299

Campbell, J. I., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory and Cognition*, *34*(3), 633–647.

Clearman, J., Klinger, V., & Szűcs, D. (2017). Visuospatial and verbal memory in mental arithmetic. *Quarterly Journal of Experimental Psychology (2006)*, *70*(9), 1837–1855. https://doi.org/10.1080/17470218.2016.1209534

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Dai, W. (2018). *The Role of Phonological and Visual Working Memory in Carry Operations or Intermediate Solutions in Complex Mental Arithmetic* [Thesis, Open Access Te Herenga Waka-Victoria University of Wellington]. https://doi.org/10.26686/wgtn.17132207.v1

De Smedt, B., Holloway, I. D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *Neuroimage*, *57*(3), 771–781.

Dehaene, S., Piazza ,Manuela, Pinel ,Philippe, & and Cohen, L. (2003). Three Parietal Circuits for Number Processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Deschuyteneer, M., De Rammelaere, S., & Fias, W. (2005). The addition of two-digit numbers: Exploring carry versus no-carry problems. *Psychology Science*, *47*(1), 74–83.

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386. https://doi.org/10.1080/09541440244000328

Emerson, R. W., & Cantlon, J. F. (2012). Early math achievement and functional connectivity in the fronto-parietal network. *Developmental Cognitive Neuroscience*, *2*, S139–S151. https://doi.org/10.1016/j.dcn.2011.11.003

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & 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. https://doi.org/10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., & De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biological Psychology*, *87*(1), 128–136. https://doi.org/10.1016/j.biopsycho.2011.02.019

Gullick, M. M., & Wolford, G. (2012). Brain systems involved in arithmetic with positive versus negative numbers. *Human Brain Mapping*, *35*(2), 539–551. https://doi.org/10.1002/hbm.22201

Imbo, I., & LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese-and Canadian-educated adults. *Memory and Cognition*, *38*(2), 176–185. https://doi.org/10.3758/MC.38.2.176

Imbo, I., Rammelaere, S. D., & Vandierendonck, A. (2005). New insights in the role of working memory in carry and borrow operations. *Psychologica Belgica*, *45*(2), Article 2. https://doi.org/10.5334/pb-45-2-101

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007a). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *Quarterly Journal of Experimental Psychology*, *60*(5), 708–731. https://doi.org/10.1080/17470210600762447

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007b). The role of working memory in carrying and borrowing. *Psychological Research*, *71*(4), 467–483. https://doi.org/10.1007/s00426-006-0044-8

Istomina, A., & Arsalidou, M. (2024). Add, subtract and multiply: Meta-analyses of brain correlates of arithmetic operations in children and adults. *Developmental Cognitive Neuroscience*, *69*, 101419. https://doi.org/10.1016/j.dcn.2024.101419

Klein, E., Huber, S., Nuerk, H.-C., & Moeller, K. (2014). Operational momentum affects eye fixation behaviour. *Quarterly Journal of Experimental Psychology*, *67*(8), 1614–1625. https://doi.org/10.1080/17470218.2014.902976

Klein, E., & Knops, A. (2023). The two-network framework of number processing: A step towards a better understanding of the neural origins of developmental dyscalculia. *Journal of Neural Transmission*, *130*(3), 253–268. https://doi.org/10.1007/s00702-022-02580-8

Klein, E., Willmes, K., Dressel, K., Domahs, F., Wood, G., Nuerk, H.-C., & Moeller, K. (2010). Categorical and continuous—Disentangling the neural correlates of the carry effect in multi-digit addition. *Behavioral and Brain Functions*, *6*(1), 70. https://doi.org/10.1186/1744-9081-6-70

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research*, *22*(3), 397–405. https://doi.org/10.1016/j.cogbrainres.2004.09.011

Lozano Iglesias, S., Spitzer, M., Strittmatter, Y., Moeller, K., & Ruiz-Garcia, M. (2025). *Towards a curriculum for neural networks to simulate symbolic arithmetic*.

Meyerhoff, H. S., Moeller, K., Debus, K., & Nuerk, H.-C. (2012). Multi-digit number processing beyond the two-digit number range: A combination of sequential and parallel processes. *Acta Psychologica*, *140*(1), 81–90. https://doi.org/10.1016/j.actpsy.2011.11.005

Moeller, K., Klein, E., & Nuerk, H.-C. (2011). Three processes underlying the carry effect in addition–Evidence from eye tracking. *British Journal of Psychology*, *102*(3), 623–645.

Montefinese, M., Turco, C., Piccione, F., & Semenza, C. (2017). Causal role of the posterior parietal cortex for two-digit mental subtraction and addition: A repetitive TMS study. *NeuroImage*, *155*, 72–81. https://doi.org/10.1016/j.neuroimage.2017.04.058

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, *16*(4), 310–322. https://doi.org/10.1016/j.learninstruc.2006.07.003

Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). Multi-digit number processing – overview, conceptual clarifications, and language influences.. In *The Oxford Handbook of*

Numerical Cognition (pp. 106–139). Oxford University Press. https://scholar.google.com/citations?view_op=view_citation&hl=zh-CN&user=pb9VhicAAAAJ&citation_for_view=pb9VhicAAAAJ:WJVC3Jt7v1AC

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, *82*(1), B25–B33.

Ortega-Martinez, A., Rogers, D., Anderson, J., Farzam, P., Gao, Y., Zimmermann, B., Yücel, M. A., & Boas, D. A. (2022). How much do time-domain functional near-infrared spectroscopy (fNIRS) moments improve estimation of brain activity over traditional fNIRS? *Neurophotonics*, *10*(1), 013504. https://doi.org/10.1117/1.NPh.10.1.013504

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, *133*(1), 188–200. https://doi.org/10.1016/j.cognition.2014.06.011

Peters, L., & De Smedt, B. (2017). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*, 265–279. https://doi.org/10.1016/j.dcn.2017.05.002

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

Riccomini, P. J. (2005). Identification and Remediation of Systematic Error Patterns in Subtraction. *Learning Disability Quarterly*, *28*(3), 233–242. https://doi.org/10.2307/1593661

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., & Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia*, *49*(9), 2592–2608. https://doi.org/10.1016/j.neuropsychologia.2011.04.035

Saban, W., Pinheiro-Chagas, P., Borra, S., & Ivry, R. B. (2024). Distinct Contributions of the Cerebellum and Basal Ganglia to Arithmetic Procedures. *Journal of Neuroscience*, *44*(2). https://doi.org/10.1523/JNEUROSCI.1482-22.2023

Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS Brain AnalyzIR Toolbox. *Algorithms*, *11*(5), 73. https://doi.org/10.3390/a11050073

Schroeder, P. A., Artemenko, C., Kosie, J. E., Cockx, H., Stute, K., Pereira, J., Klein, F., & Mehler, D. M. A. (2023). Using preregistration as a tool for transparent fNIRS study design. *Neurophotonics*, *10*(2), 023515. https://doi.org/10.1117/1.NPh.10.2.023515

Selter, C. (2001). Addition and Subtraction of Three-digit Numbers: German Elementary Children's Success, Methods and Strategies. *Educational Studies in Mathematics*, *47*(2), 145–173. https://doi.org/10.1023/A:1014521221809

Sokolowski, H. M., Matejko, A. A., & Ansari, D. (2023). The role of the angular gyrus in arithmetic processing: A literature review. *Brain Structure & Function*, *228*(1), 293–304. https://doi.org/10.1007/s00429-022-02594-8

Soltanlou, M., Artemenko, C., Ehlis, A.-C., Huber, S., Fallgatter, A. J., Dresler, T., & Nuerk, H.-C. (2018). Reduction but no shift in brain activation after arithmetic learning in children: A simultaneous fNIRS-EEG study. *Scientific Reports*, *8*(1), 1707. https://doi.org/10.1038/s41598-018-20007-x

Trbovich, P. L., & LeFevre, J.-A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, *31*(5), 738–745. https://doi.org/10.3758/bf03196112

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM

Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Uittenhove, K., Burger, L., Taconnat, L., & Lemaire, P. (2015). Sequential difficulty effects during execution of memory strategies in young and older adults. *Memory*, *23*(6), 806–816. https://doi.org/10.1080/09658211.2014.928730

Verner, M., Herrmann, M. J., Troche, S. J., Roebers, C. M., & Rammsayer, T. H. (2013). Cortical oxygen consumption in mental arithmetic as a function of task difficulty: A near-infrared spectroscopy approach. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00217

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

Yao, X., Artemenko, C., He, Y., & Nuerk, H.-C. (2025). Arithmetic is not arithmetic: Paradigm matters for arithmetic effects. *Cognition*, *256*, 106060.

Yi-Rong, N., Si-Yun, S., Zhou-Yi, G., Si-Run, L., Yun, B., Song-Hao, L., & Chan, W. Y. (2011). Dissociated brain organization for two-digit addition and subtraction: An fMRI investigation. *Brain Research Bulletin*, *86*(5), 395–402. https://doi.org/10.1016/j.brainresbull.2011.08.016

Zhang, J., Zhao, N., & Kong, Q. P. (2019). The relationship between math anxiety and math performance: A meta-analytic investigation. *Frontiers in Psychology*, *10*, 458192. https://doi.org/10.3389/fpsyg.2019.01613

A4.2 Supplementary Materials for Study 4

These supplementary materials belong to the following publication:

Yao, X., Barth, B., & Artemenko, C. When arithmetic gets complex: fNIRS evidence for fronto-parietal activation in multi-digit arithmetic. *npj* Science of Learning (under review)

Supplementary Materials

Title: When Arithmetic Gets Complex: fNIRS Evidence for Fronto-parietal Activation in Multi-digit Arithmetic

Authors: Xinru Yao, Beatrix Barth, Christina Artemenko

Table S1. Descriptive information on accuracy and reaction time and carry/borrow effects

	addition					**subtraction**				
	complex_0	complex_1	complex_2	carry effect$_{0 \text{ to } 1}$	carry effect$_{1 \text{ to } 2}$	complex_0	complex_1	complex_2	borrow effect$_{0 \text{ to } 1}$	borrow effect$_{1 \text{ to } 2}$
accuracy	0.96 ± 0.05	0.92 ± 0.07	0.87 ± 0.10	0.04 ± 0.07	0.05 ± 0.09	0.93 ± 0.07	0.88 ± 0.09	0.82 ± 0.14	0.04 ± 0.08	0.06 ± 0.10
reaction time	5.11 ± 1.20	7.09 ± 1.82	8.33 ± 2.25	1.99 ± 0.96	1.24 ± 0.86	6.50 ± 1.77	9.42 ± 2.89	11.26 ± 3.23	2.93 ± 1.57	1.83 ± 1.51

Table S2. Descriptive information on mean amplitude of HbO and HbR.

Channel	HbO			HbR		
	Complexity_0	Complexity_1	Complexity_2	Complexity_0	Complexity_1	Complexity_2
LP 1	0.090 (0.113)	0.118 (0.125)	0.132 (0.126)	0.004 (0.091)	-0.036 (0.113)	-0.051 (0.131)
LP 2	0.111 (0.116)	0.128 (0.121)	0.116 (0.111)	-0.005 (0.101)	-0.014 (0.088)	-0.021 (0.104)
LP 3	0.078 (0.117)	0.094 (0.123)	0.088 (0.127)	0.056 (0.089)	0.060 (0.090)	0.043 (0.116)
LF 4	0.098 (0.106)	0.131 (0.128)	0.141 (0.142)	-0.029 (0.109)	-0.051 (0.134)	-0.070 (0.149)
LF 5	0.079 (0.092)	0.125 (0.104)	0.140 (0.122)	-0.022 (0.099)	-0.061 (0.118)	-0.080 (0.144)
LF 6	0.068 (0.109)	0.087 (0.105)	0.092 (0.114)	0.067 (0.086)	0.051 (0.093)	0.056 (0.107)
LF 7	0.068 (0.101)	0.090 (0.101)	0.080 (0.122)	0.035 (0.085)	0.010 (0.105)	-0.001 (0.131)
LF 8	0.048 (0.079)	0.088 (0.076)	0.108 (0.089)	0.013 (0.069)	0.002 (0.076)	-0.016 (0.106)
RP 9	0.087 (0.136)	0.107 (0.135)	0.121 (0.129)	0.010 (0.096)	-0.003 (0.086)	-0.014 (0.110)
RP 10	0.113 (0.117)	0.123 (0.121)	0.121 (0.113)	0.027 (0.080)	0.021 (0.060)	0.009 (0.075)
RP 11	0.092 (0.114)	0.106 (0.118)	0.109 (0.129)	0.044 (0.116)	0.039 (0.112)	0.034 (0.103)
RF 12	0.074 (0.101)	0.088 (0.100)	0.089 (0.087)	-0.007 (0.125)	-0.017 (0.126)	-0.021 (0.117)
RF 13	0.071 (0.100)	0.096 (0.099)	0.125 (0.101)	-0.013 (0.091)	-0.041 (0.095)	-0.067 (0.114)
RF 14	0.062 (0.098)	0.090 (0.092)	0.095 (0.084)	0.055 (0.087)	0.057 (0.077)	0.043 (0.080)
RF 15	0.068 (0.104)	0.094 (0.100)	0.092 (0.111)	0.009 (0.086)	0.012 (0.107)	-0.002 (0.139)
RF 16	0.044 (0.089)	0.067 (0.104)	0.087 (0.112)	0.018 (0.077)	0.027 (0.092)	0.006 (0.099)

Notes. Mean (SD) of Mean Amplitude for each channel across the three complexity levels are shown separately for HbO and HbR.

Table S3. Two-way ANOVA for β-weights for HbO

HbO			Complexity			Operation			Complexity * Operation		
Channel	**Brain region**	**MNI (x, y, z)**	***F***	***p***	***q***	***F***	***p***	***q***	***F***	***p***	***q***
LP 1	left SMG	-56.851, -46.507, 50.825	0.31	.734	.988	2.34	.127	.988	0.70	.497	.988
LP 2	left IPS	-42.484, -56.425, 59.293	0.11	.894	.988	0.68	.412	.988	0.15	.861	.988
LP 3	left AG	-45.161, -69.940, 48.359	1.54	.216	.988	6.60	.011	.618	0.30	.743	.988
LF 4	left MFG	-52.981, 9.441, 43.288	1.34	.262	.988	2.87	.091	.972	0.41	.661	.988
LF 5	left IFG	-54.989, 21.577, 28.224	1.54	.216	.988	3.68	.056	.937	0.44	.642	.988
LF 6	left MFG	-40.683, 22.891, 50.628	4.00	.019	.618	3.52	.062	.937	0.76	.469	.988
LF 7	left MFG	-43.559, 36.844, 35.329	1.90	.152	.988	0.73	.394	.988	1.03	.357	.988
LF 8	left IFG	-54.192, 35.192, 10.404	2.85	.060	.937	1.38	.241	.988	1.39	.250	.988
RP 9	right SMG	56.900, -46.866, 51.037	0.56	.570	.988	1.55	.215	.988	0.40	.669	.988
RP 10	right IPS	41.785, -56.609, 59.831	1.00	.371	.988	2.06	.152	.988	0.36	.695	.988
RP 11	right AG	43.703, -69.098,	1.01	.366	.988	1.25	.265	.988	0.17	.840	.988

		48.517									
RF 12	right MFG	54.448, 8.156, 43.809	1.11	.330	.988	0.03	.868	.988	0.11	.894	.988
RF 13	right IFG	56.431, 21.530, 28.382	1.48	.230	.988	1.77	.184	.988	0.10	.909	.988
RF 14	right MFG	41.653, 22.829, 50.474	4.39	.013	.618	2.64	.105	.988	0.21	.807	.988
RF 15	right MFG	45.204, 35.666, 36.000	1.30	.274	.988	0.00	.994	.998	0.87	.421	.988
RF 16	right IFG	56.186, 34.089, 11.569	0.97	.381	.988	0.31	.576	.988	0.75	.471	.988

Notes. No significant main effects or interactions were observed in any channels for β-weights for HbO. LP – left parietal, LF – left frontal, RP – right parietal, RF – right frontal. MNI coordinates indicate the correspondence of fNIRS channels to the underlying cortical areas. *q* value represents the statistics after FDR correction.

Table S4. Two-way ANOVA for β-weights for HbR

HbO			**Complexity**			**Operation**			**Complexity * Operation**		
Channel	**Brain region**	**MNI (x, y, z)**	***F***	***p***	***q***	***F***	***p***	***q***	***F***	***p***	***q***
LP 1	left SMG	-56.851, -46.507, 50.825	0.13	.876	.988	0.02	.897	.988	0.06	.937	.988
LP 2	left IPS	-42.484, -56.425, 59.293	1.01	.367	.988	0.01	.910	.988	0.20	.818	.988
LP 3	left AG	-45.161, -69.940, 48.359	0.87	.421	.988	0.05	.829	.988	0.04	.958	.989
LF 4	left MFG	-52.981, 9.441, 43.288	0.15	.861	.988	0.05	.829	.988	0.14	.871	.988
LF 5	left IFG	-54.989, 21.577, 28.224	0.24	.783	.988	0.44	.509	.988	0.21	.811	.988
LF 6	left MFG	-40.683, 22.891, 50.628	1.23	.293	.988	1.01	.316	.988	0.34	.714	.988
LF 7	left MFG	-43.559, 36.844, 35.329	0.54	.584	.988	0.12	.724	.988	0.27	.761	.988
LF 8	left IFG	-54.192, 35.192, 10.404	2.71	.068	.937	1.53	.217	.988	0.07	.928	.988
RP 9	right SMG	56.900, -46.866, 51.037	0.08	.926	.988	0.79	.375	.988	0.41	.665	.988
RP 10	right IPS	41.785, -56.609, 59.831	2.43	.090	.972	0.54	.465	.988	0.37	.688	.988
RP 11	right AG	43.703, -69.098, 48.517	0.54	.583	.988	1.47	.227	.988	0.01	.988	.998

RF 12	right MFG	54.448, 8.156, 43.809	0.30	.744	.988	0.53	.468	.988	0.17	.845	.988
RF 13	right IFG	56.431, 21.530, 28.382	0.16	.851	.988	0.45	.503	.988	0.00	.998	.998
RF 14	right MFG	41.653, 22.829, 50.474	0.10	.903	.988	0.25	.620	.988	0.06	.946	.988
RF 15	right MFG	45.204, 35.666, 36.000	0.07	.929	.988	0.06	.801	.988	0.10	.905	.988
RF 16	right IFG	56.186, 34.089, 11.569	0.45	.640	.988	0.31	.580	.988	0.38	.681	.988

Notes. No significant main effects or interactions were observed in any channels for β-weights for HbR. LP – left parietal, LF – left frontal, RP – right parietal, RF – right frontal. MNI coordinates indicate the correspondence of fNIRS channels to the underlying cortical areas. *q* value represents the statistics after FDR correction.

Table S5. Two-way ANOVA for mean amplitude for HbO

HbO			**Complexity**			**Operation**			**Complexity * Operation**		
Channel	**Brain region**	**MNI (x, y, z)**	***F***	***p***	***q***	***F***	***p***	***q***	***F***	***p***	***q***
LP 1	left SMG	-56.851, -46.507, 50.825	4.23	.016	**.048**	0.40	.527	.815	0.56	.570	.815
LP 2	left IPS	-42.484, -56.425, 59.293	9.24	.000	**.000**	0.09	.763	.891	0.51	.560	.827
LP 3	left AG	-45.161, -69.940, 48.359	28.23	.000	**.000**	1.42	.235	.562	0.28	.753	.891
LF 4	left MFG	-52.981, 9.441, 43.288	20.03	.000	**.000**	0.26	.611	.827	0.93	.397	.750
LF 5	left IFG	-54.989, 21.577, 28.224	16.61	.000	**.000**	0.08	.782	.891	0.09	.912	.979
LF 6	left MFG	-40.683, 22.891, 50.628	36.18	.000	**.000**	0.10	.751	.891	0.26	.767	.891
LF 7	left MFG	-43.559, 36.844, 35.329	38.04	.000	**.000**	0.00	.980	.990	1.28	.279	.622
LF 8	left IFG	-54.192, 35.192, 10.404	12.11	.000	**.000**	0.00	.958	.979	0.58	.557	.815
RP 9	right SMG	56.900, -46.866, 51.037	2.01	.136	.383	0.09	.760	.891	0.85	.428	.775
RP 10	right IPS	41.785, -56.609, 59.831	5.86	.003	**.010**	0.03	.858	.946	0.07	.923	.979
RP 11	right AG	43.703, -69.098, 48.517	27.10	.000	**.000**	0.00	.957	.979	0.87	.421	.775

RF 12	right MFG	54.448, 8.156, 43.809	36.34	.000	**.000**	0.80	.373	.740	0.71	.494	.805
RF 13	right IFG	56.431, 21.530, 28.382	17.79	.000	**.000**	0.00	.992	.992	0.65	.521	.815
RF 14	right MFG	41.653, 22.829, 50.474	29.32	.000	**.000**	0.22	.638	.827	1.51	.222	.561
RF 15	right MFG	45.204, 35.666, 36.000	39.23	.000	**.000**	0.56	.453	.777	1.10	.335	.670
RF 16	right IFG	56.186, 34.089, 11.569	9.11	.000	**.001**	0.06	.780	.893	0.56	.574	.815

Notes. The main effect of complexity was significant for HbO in all channels except for the right SMG. LP – left parietal, LF – left frontal, RP – right parietal, RF – right frontal. MNI coordinates indicate the correspondence of fNIRS channels to the underlying cortical areas. *q* value represents the statistics after FDR correction.

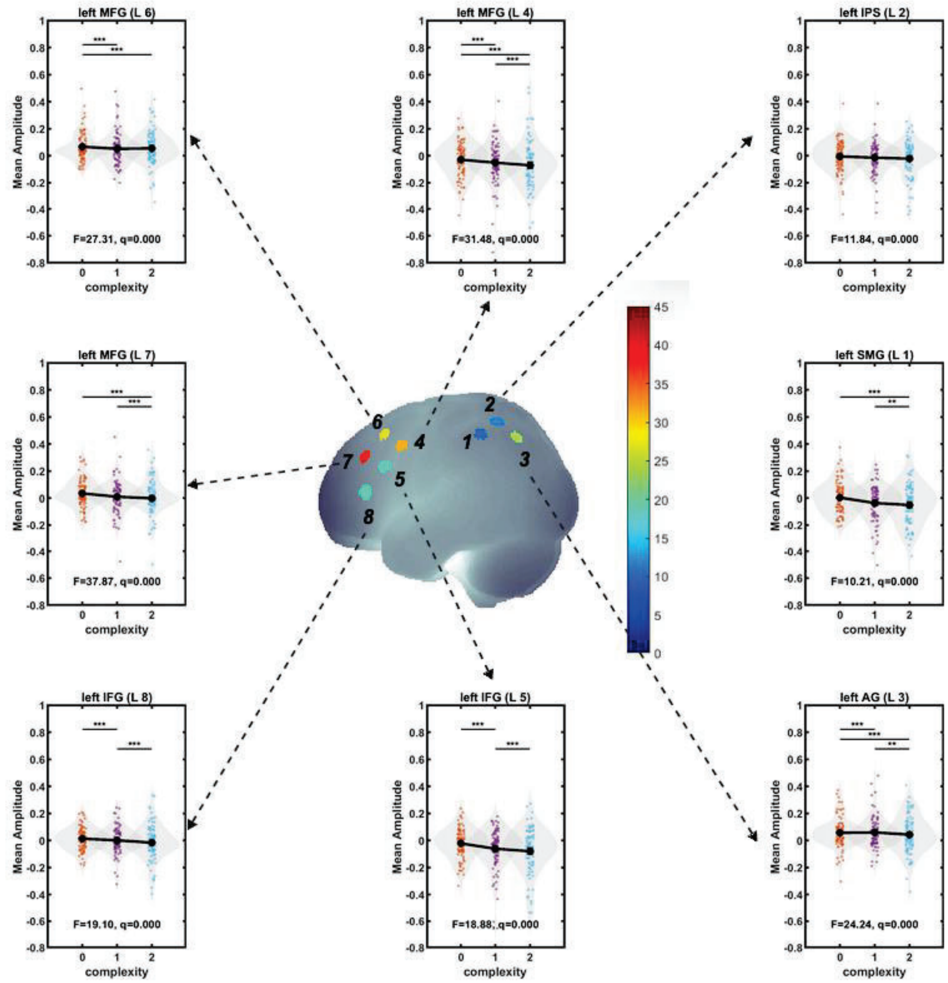
Table S6. Two-way ANOVA for mean amplitude for HbR

HbR			Complexity			Operation			Complexity * Operation		
Channel	**Brain region**	**MNI (x, y, z)**	***F***	***p***	***q***	***F***	***p***	***q***	***F***	***p***	***q***
LP 1	left SMG	-56.851, -46.507, 50.825	10.21	.000	**.000**	0.37	.545	.815	1.88	.155	.424
LP 2	left IPS	-42.484, -56.425, 59.293	11.84	.000	**.000**	0.56	.454	.777	0.98	.378	.740
LP 3	left AG	-45.161, -69.940, 48.359	24.24	.000	**.000**	0.23	.635	.827	0.78	.461	.777
LF 4	left MFG	-52.981, 9.441, 43.288	31.48	.000	**.000**	0.20	.655	.838	0.24	.789	.891
LF 5	left IFG	-54.989, 21.577, 28.224	18.88	.000	**.000**	0.24	.626	.827	0.24	.787	.891
LF 6	left MFG	-40.683, 22.891, 50.628	27.31	.000	**.000**	0.31	.577	.815	1.36	.257	.588
LF 7	left MFG	-43.559, 36.844, 35.329	37.87	.000	**.000**	0.97	.326	.670	1.10	.335	.670
LF 8	left IFG	-54.192, 35.192, 10.404	19.10	.000	**.000**	1.39	.240	.562	0.78	.458	.777
RP 9	right SMG	56.900, -46.866, 51.037	5.97	.003	**.010**	0.07	.789	.891	0.59	.554	.815
RP 10	right IPS	41.785, -56.609, 59.831	24.26	.000	**.000**	2.31	.130	.378	0.92	.398	.750

RP 11	right AG	43.703, -69.098, 48.517	25.44	.000	**.000**	0.25	.620	.827	0.62	.540	.815
RF 12	right MFG	54.448, 8.156, 43.809	38.86	.000	**.000**	0.00	.959	.979	0.72	.490	.805
RF 13	right IFG	56.431, 21.530, 28.382	23.73	.000	**.000**	2.66	.104	.312	1.48	.229	.562
RF 14	right MFG	41.653, 22.829, 50.474	16.30	.000	**.000**	0.07	.787	.891	0.07	.932	.979
RF 15	right MFG	45.204, 35.666, 36.000	41.66	.000	**.000**	0.03	.868	.946	0.99	.375	.740
RF 16	right IFG	56.186, 34.089, 11.569	9.04	.000	**.010**	1.59	.208	.540	1.67	.190	.507

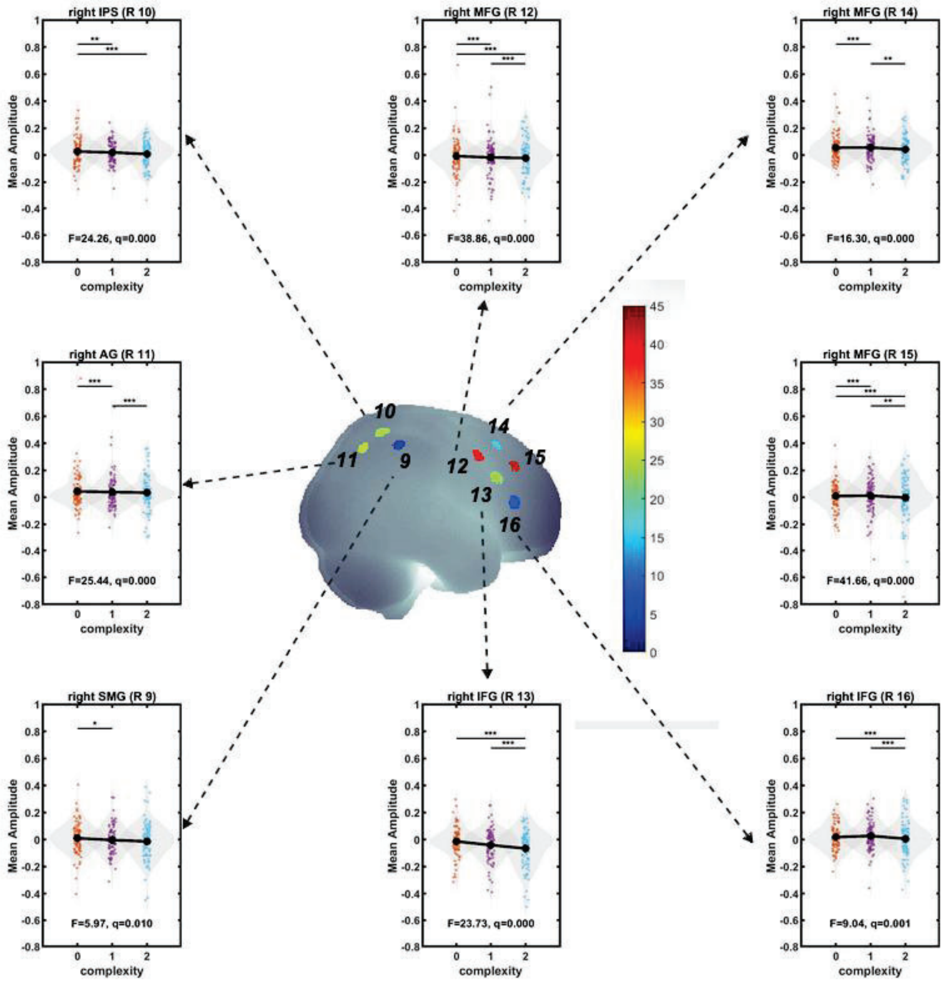
Notes. The main effect of complexity was significant for HbR in all frontal and parietal channels. LP – left parietal, LF – left frontal, RP – right parietal, RF – right frontal. MNI coordinates indicate the correspondence of fNIRS channels to the underlying cortical areas. *q* value represents the statistics after FDR correction.

Figure S1. Fronto-parietal activation for complexity effects in the left hemisphere (HbR mean amplitude)



Notes. Colors on the cortical map denote the F statistic of the main effect of complexity (0/1/2) from two-way ANOVAs on HbR mean amplitude at each left-hemisphere channel (warmer colors = larger F). Violin plots show participant distributions; black dots indicate means connected by lines; asterisks mark Bonferroni-adjusted post-hoc pairwise contrasts (*** $p < .001$, ** $p < .01$, * $p < .05$). Descriptive means ± SD for all channels and levels are provided in Table S2.

Figure S2. Fronto-parietal activation for complexity effects in the right hemisphere (HbR mean amplitude)



Notes. Colors on the cortical map denote the F statistic of the main effect of complexity (0/1/2) from two-way ANOVAs on HbR mean amplitude at each right-hemisphere channel (warmer colors = larger F). Violin plots show participant distributions; black dots indicate means connected by lines; asterisks mark Bonferroni-adjusted post-hoc pairwise contrasts (*** $p < .001$, ** $p < .01$, * $p < .05$). Descriptive means ± SD for all channels and levels are provided in Table S2.

Table S7. Correlations between working memory and carry/borrow effect on behavioral level (accuracy and reaction time)

	$\Delta ACC_{0\ to\ 1}$	$\Delta ACC_{1\ to\ 2}$	$\Delta RT_{0\ to\ 1}$	$\Delta RT_{1\ to\ 2}$
verbal short-term memory	0.01	-0.44 **	-0.06	-0.23
verbal working memory	-0.19	-0.31 *	-0.20	-0.17
visuospatial short-term memory	-0.04	0.04	-0.19	-0.02
visuospatial working memory	-0.30*	-0.10	-0.08	-0.24

***Notes*.** Pearson's correlation values were presented in the table. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S8. Correlations between working memory and carry effect on neural level in HbO (mean amplitude)

Ch	Brain regions	Verbal short-term memory		Verbal working memory		Visuospatial short-term memory		Visuospatial working memory	
		$\Delta HbO_{0\ to\ 1}$	$\Delta HbO_{1\ to\ 2}$	$\Delta HbO_{0\ to\ 1}$	$\Delta HbO_{1\ to\ 2}$	$\Delta HbO_{0\ to\ 1}$	$\Delta HbO_{1\ to\ 2}$	$\Delta HbO_{0\ to\ 1}$	$\Delta HbO_{1\ to\ 2}$
LP1	left SMG	-0.04	-0.23	-0.10	-0.21	0.12	-0.28	0.11	-0.04
LP2	left IPS	-0.19	-0.12	-0.20	-0.03	-0.04	-0.17	0.01	0.06
LP3	left AG	-0.07	-0.18	-0.15	-0.04	0.07	-0.27	0.06	-0.08
LF4	left MFG	0.10	-0.06	0.16	-0.08	0.10	-0.20	0.04	-0.13
LF5	left IFG	-0.04	-0.08	-0.10	0.06	0.04	-0.22	0.14	-0.22
LF6	left MFG	0.02	-0.32 *	-0.03	-0.16	0.02	-0.24	0.01	-0.18
LF7	left MFG	-0.00	-0.23	0.02	-0.07	-0.06	-0.22	0.03	-0.23
LF8	left IFG	0.00	-0.12	0.08	0.01	0.02	-0.24	0.13	-0.25
RP9	right SMG	0.01	-0.09	-0.11	-0.08	-0.04	-0.30 *	-0.05	-0.11
RP10	right IPS	-0.05	-0.15	-0.13	-0.04	0.00	-0.27	-0.05	-0.15
RP11	right AG	-0.03	-0.22	-0.16	-0.17	0.06	-0.31 *	0.09	-0.06
RF12	right MFG	0.14	-0.12	0.00	-0.13	0.18	-0.12	0.08	-0.02
RF13	right IFG	0.17	0.00	0.09	0.04	0.19	-0.04	0.09	0.02
RF14	right MFG	-0.08	-0.30 *	-0.11	-0.22	0.22	-0.20	0.14	-0.18
RF15	right MFG	0.07	-0.14	0.02	-0.21	0.14	-0.09	-0.05	0.07
RF16	right IFG	-0.01	-0.08	0.17	-0.03	-0.15	-0.24	-0.04	-0.10

Notes. Pearson's correlation values were presented in the table. *** $p < .001$, ** $p < .01$, * $p < .05$.

Table S9. Correlation between working memory and carry effect on neural level in HbR (mean amplitude)

Ch	Brain regions	Verbal short-term memory		Verbal working memory		Visuospatial short-term memory		Visuospatial working memory	
		$\Delta HbR_{0\ to\ 1}$	$\Delta HbR_{1\ to\ 2}$	$\Delta HbR_{0\ to\ 1}$	$\Delta HbR_{1\ to\ 2}$	$\Delta HbR_{0\ to\ 1}$	$\Delta HbR_{1\ to\ 2}$	$\Delta HbR_{0\ to\ 1}$	$\Delta HbR_{1\ to\ 2}$
LP1	left SMG	0.32 *	-0.25	0.27	-0.32 *	0.16	-0.12	0.23	-0.03
LP2	left IPS	0.08	-0.00	0.01	0.16	0.05	-0.02	0.05	-0.01
LP3	left AG	0.09	-0.17	0.11	-0.11	0.16	-0.17	0.04	-0.21
LF4	left MFG	0.14	-0.03	0.09	-0.14	-0.16	-0.01	-0.11	0.11
LF5	left IFG	0.11	0.05	-0.10	-0.03	0.18	-0.14	0.35 *	-0.26
LF6	left MFG	-0.05	0.08	0.01	-0.08	-0.21	-0.02	-0.05	-0.20
LF7	left MFG	0.10	-0.01	-0.09	-0.15	0.07	0.07	-0.09	-0.27
LF8	left IFG	0.06	0.20	-0.05	0.15	-0.17	-0.01	0.14	-0.30 *
RP9	right SMG	0.18	-0.15	0.00	-0.42 **	-0.04	-0.31 *	0.01	-0.07
RP10	right IPS	0.01	0.19	0.03	-0.03	-0.20	-0.03	-0.09	-0.08
RP11	right AG	0.05	-0.01	0.02	-0.11	-0.14	-0.14	-0.01	-0.03
RF12	right MFG	0.25	0.03	0.10	-0.26	-0.17	0.05	-0.13	0.23
RF13	right IFG	-0.03	0.15	-0.17	-0.16	0.02	0.09	0.16	0.06
RF14	right MFG	-0.01	-0.04	-0.20	-0.06	-0.33 *	0.08	-0.08	0.06
RF15	right MFG	0.25	-0.07	0.05	-0.25	0.15	0.10	-0.05	0.11
RF16	right IFG	-0.12	-0.05	-0.18	-0.21	-0.26	0.00	-0.06	-0.13

***Notes*.** Pearson's correlation values were presented in the table. *** $p < .001$, ** $p < .01$, * $p < .05$.