Differential Switch Costs in Typically Achieving Children and Children with Mathematical Difficulties

Darius Endlich, Wolfgang Lenhard, Peter Marx, and Tobias Richter

University of Würzburg, Faculty of Human Sciences, Department of Psychology IV, Germany

Accepted for publication in

***Journal of Learning Disabilities* (2023)**

Author Note

Darius Endlich https://orcid.org/0000-0002-5647-7671

Wolfgang Lenhard https://orcid.org/0000-0002-8184-6889

Peter Marx https://orcid.org/0000-0002-1671-6148

Tobias Richter https://orcid.org/0000-0002-0467-9044

The data and analysis scripts are available at the OSF repository (https://osf.io/f8pzk/?view_only=cd81fff16b9348f5995980d13cc753a6). We have no conflicts of interest to disclose. Our research was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF), Grant 01GJ1710B.

Correspondence concerning this article should be addressed to Darius Endlich, University of Würzburg, Röntgenring 10, 97070 Würzburg, Germany. Email: darius.endlich@uni-wuerzburg.de

Abstract

Children with mathematical difficulties need to spend more time than typically achieving children on solving even simple equations. Since these tasks already require a larger share of their cognitive resources, additional demands imposed by the need to switch between tasks may lead to a greater decline of performance in children with mathematical difficulties. We explored differential task switch costs with respect to switching between addition versus subtraction with a tablet-based arithmetic verification task and additional standardized tests in elementary school children in Grades 1 to 4. Two independent studies were conducted. In Study 1, we assessed the validity of a newly constructed tablet-based arithmetic verification task in a controlled classroom-setting (n = 165). Then, effects of switching between different types of arithmetic operations on accuracy and response latency were analyzed through Generalized Linear Mixed Models (GLMM) in an online-based testing (Study 2; n = 3,409). Children with mathematical difficulties needed more time and worked less accurately overall. They also exhibited a stronger performance decline when working in a task switching condition, when working on subtraction (vs. addition) items and in operations with two-digit (vs. one-digit) operations. These results underline the value of process data in the context of assessing mathematical difficulties.

Keywords: mathematical difficulties, arithmetic verification, tablet-based testing

Differential Switch Costs in Typically Achieving Children and Children with Mathematical Difficulties

Deficits for children with mathematical difficulties appear quite early in elementary school, prior to Grade 3 (Jordan et al., 2003). At the end of elementary school, around 20% of the children fall short of the minimum standard in mathematics in Germany (Stanat et al., 2022). Without targeted interventions, these difficulties often remain stable until adolescence (and beyond) for most of the children (Shalev et al., 2005). Low mathematical skills in elementary school increase the risk for low socio-economic status in adulthood (Ritchie & Bates, 2013) as well as early school leaving, unemployment, adult mental health, and juvenile delinquency (Parsons & Bynner, 2005). Obviously, an early identification of children with mathematical difficulties is crucial in order to intervene as early as possible and reduce the life problems typically associated with deficits in mathematics.

Deficits in foundational arithmetic fact knowledge are frequently observed in children with mathematical difficulties (Geary, 1993; Jordan & Hanich, 2003). To clarify, arithmetic, in this context, refers to the part of mathematics that deals with numbers and basic operations such as addition, subtraction, multiplication, and division. Adequate knowledge of foundational arithmetic facts is crucial for efficient problem-solving in calculations. This includes, for instance, the ability to recall addition and subtraction facts within the number range of 1 to 20 from long-term memory (Ashcraft, 1982; Siegler & Shrager, 1984). The availability of foundational arithmetic fact knowledge affects the cognitive load while working on arithmetic tasks. Retrieval of arithmetic fact knowledge requires less resources than using counting strategies (Grube, 2006; Kaye, 1986). In children with mathematical difficulties, both the availability of factual knowledge and speed of retrieval are restricted (Busch et al., 2013). Therefore, additional demands imposed by task-switching conditions (in our study operationalized as the need to switch between addition and subtraction tasks) should place an excessive cognitive load on children with mathematical difficulties, leading to an additional loss of performance which is greater than in normally developing children.

Foundational arithmetic competencies refer to basic mathematical skills and knowledge required for success in higher-level mathematics (e.g. understanding numbers and mathematical concepts). These competencies are strongly associated with arithmetic fluency tasks, such as solving as many arithmetic tasks as possible in a limited time (J. I. D. Campbell & Tarling, 1996; Dewi et al., 2021). In our studies, we investigated differential switch costs in typically achieving children and children with mathematical difficulties with a newly constructed computerized arithmetic verification task. While prevalence rates for dyscalculia–a persistent difficulty in learning arithmetic–vary between 2% and 7% (Devine et al., 2013; Rapin, 2016), the number of children who do not learn basic competencies during primary education is significantly higher at 15% to 20% in North America and Europe (UNESCO, 2017). The majority of mathematical difficulties can be attributed to various aspects, for instance environmental factors such as the extent of education or early learning environment. In our studies, we examined children with mathematical difficulties, irrespective of the cause of difficulties, and defined children who achieved below-average scores in mathematical tests compared to the reference group (≤ 16th percentile) as children with mathematical difficulties.

Development of Foundational Arithmetic Fact Knowledge and Arithmetic Fact Fluency

Arithmetic concepts and, consequently, addition and subtraction skills, develop throughout elementary school, basically following the curriculum – for instance, addition precedes subtraction (Rubinstein et al., 2001). Addition and subtraction both require understanding of additive composition and part-whole relations (Butterworth, 2005; Nunes et al., 2016). Any subtraction problem can be transformed into an addition problem (e.g., 2 + 4 = 6, 6 – 4 = 2, 6 – 2 = 4). Thus, these operations are complementary both procedurally and conceptually (Robinson, 2017; Robinson & Dubé, 2009, 2012).

Similar to the dual-route model of reading (Coltheart, 1978), two possible ways of calculating have been proposed for arithmetic (Amalric & Dehaene, 2019; Dehaene & Cohen, 1997). In the first, semantically mediated route, calculation strategies are necessary to solve the task. Having solved the same calculation task often enough by strategies such as counting, children store numbers and operators (e.g., "+" or "-") together with the result as foundational arithmetic fact knowledge in long-term memory (Dehaene & Cohen, 1995). Retrieving foundational arithmetic fact knowledge directly from long-term memory represents the second route that is far more efficient than the first route.

We examined this efficiency in the present study by means of fluency tasks. Fluency is often used to designate the smooth and effortless production of speech and pronunciation (Chambers, 1997) but the term is used in the field of arithmetic as well in the sense of processing fluency (Vanbinst et al., 2015). Arithmetic fact fluency refers to the automatic retrieval of simple single-digit facts from long-term memory (Zaunmüller et al., 2009). It is strongly associated with overall mathematical achievements, especially during elementary school (Nunes et al., 2012). Jordan et al. (2003) showed that children with difficulties in arithmetic fact fluency in Grade 2 had a higher risk for lower mathematical performance in later school years.

In the present study, we focused on the development across elementary school. Therefore, we should note that arithmetic fact fluency develops in several stages. Young children usually concentrate on lower-level strategies, such as counting. At this developmental stage, children often use specific external representations such as fingers or objects to manipulate quantities and perform simple arithmetic tasks (Crollen & Noël, 2015; Geary et al., 1991). With increasing conceptual and procedural knowledge about numbers, higher-order strategies evolve. At this point, children can decompose a presented arithmetic problem into easier and more familiar tasks (e.g., "8 + 5" is decomposed to "(5 + 5 = 10) + 3 = 13"; Laski et al., 2013). This counting and decomposition strategy reduces cognitive load by segmenting the problem into easier substeps, solvable through retrieval from long-term memory. As a result of an arithmetic intervention, children with mathematical difficulties increase their use of decomposition as preferred strategy and decrease their use of counting-based strategies – which are their most common strategies before the intervention (Koponen et al., 2018). Finally, the retrieval of arithmetic fact knowledge becomes more efficient, reflected by higher accuracy rates and faster processing speed (Mabbott & Bisanz, 2003).

When children begin to engage with numbers and arithmetical problems, e.g., addition and subtraction tasks, they use counting procedures as their dominant strategy (Bagnoud et al., 2021; Baroody et al., 2006). However, it is not completely clear how the further development of arithmetic skills proceeds. Two theoretical views on children's progress towards proficient processing of arithmetic problems can be contrasted. *Retrieval models* suggest that counting procedures are more and more replaced by memory retrieval (e.g., Chen & Campbell, 2018; Siegler, 1996). In contrast, simple arithmetic problems can also be solved by using rules (e.g. $N + 0 = N$ rule) and heuristics (children may reason out the products of near-ties by recalling the product of the more easily recalled tie – e.g., 7 x 7 is 49, 8 x 7 is one more seven, so its product is 49 + 7, or 56; Baroody, 1983, 1984, 1994). According to the *automated counting procedure theory*, the development of strategy consists of an acceleration of counting procedures until automatization (e.g., Barrouillet & Thevenot, 2013; Fayol & Thevenot, 2012; Mathieu et al., 2016; Thevenot et al., 2016; Uittenhove et al., 2016).

Logan (1988) described the shift from counting to retrieval in terms of his *instance theory of automatization*. According to this theory, each time children work on an arithmetic task, a single memory trace is created containing the task and the result. These traces are finally stored in long-term memory. When practice continues, more and more memory traces are created using algorithm-based procedures, resulting in a higher probability of using memory retrieval. In the end, children shift completely from counting to retrieval.

The Role of Working Memory and Development of Calculation Strategies

Deficits in working memory are strongly associated with mathematical difficulties (Friso-van den Bos et al., 2013; Raghubar et al., 2010; Schuchardt & Mähler, 2010). All main components of working memory as distinguished in Baddeley's (1986) model of working memory are relevant for completing arithmetic tasks: The central executive guides attention and ensures updating of information (Andersson & Lyxell, 2007; van der Sluis et al., 2004), the phonological loop serves as a storage system for the storage for input, intermediate results and the solution (De Weerdt et al., 2013) and the visuospatial sketchpad serves for visualization and magnitude estimation (D'Amico & Guarnera, 2005; Wilson & Swanson, 2001).

Being faced with a mathematical problem, the central executive controls calculation strategies such as retrieval of solutions or counting strategies. Retrieval of arithmetic fact knowledge demands less resources than using counting strategies (Kaye, 1986). Given that the capacity of working memory is limited, an overload of working memory may result when task difficulty increases (e.g., solving another task simultaneously), resulting in higher error rates or increased processing time (Busch et al., 2013).

One way to reduce the load on working memory is to use more efficient strategies for solving arithmetic problems. Over the course of elementary school, calculation strategies develop continuously (Siegler, 1991; Widaman et al., 1992). From counting with fingers or objects to verbally counting numbers and, finally, automatic representation and manipulation of magnitudes as well as retrieval of fact knowledge and mathematical derivation, calculation strategies become more and more efficient – although children might use different strategies at the same stage of development (and not always the most efficient one; Carpenter & Moser, 1984). Some arithmetic facts seem to be added faster to the fact knowledge, such as tasks with identical addends or tasks resulting in 10 (Gaidoschik, 2010). Generally, large interindividual differences exist in the development of calculation strategies: Whereas some children already use fact retrieval in Grade 2 (Widaman et al., 1992) or even as early as Grade 1 (Carpenter &

Moser, 1984), some adults still use counting strategies at least in some cases (J. I. D. Campbell & Fugelsang, 2001).

For number sets above 10 and for subtraction tasks, other calculation strategies must be considered or adapted (Geary et al., 1993; Siegler, 1991), even if counting is still possible. These additional strategies address especially the ten transition, e.g. the isolated handling of single-digit and multi-digit numbers. Children with mathematical difficulties face specific problems using arithmetic strategies and shifting between arithmetic strategies (Rourke, 1993; van der Sluis et al., 2004). These difficulties stem from conceptual problems in understanding underlying magnitude representations of numbers and the storage and retrieval of factual knowledge from long-term memory. For example, children with mathematical difficulties often fail to represent numbers by the underlying quantity and consequently face problems with part-whole-relations (Krajewski & Schneider, 2009). These concepts, which are fundamental for overcoming counting strategies, seem to be poorly represented and poorly interconnected in long-term memory. Consequently, associations between arithmetic tasks and their solutions cannot be established sufficiently (Siegler & Shipley, 1995). Moreover, a hyper-sensibility for similarities has been described as well, meaning that children with mathematical difficulties mix up different tasks with identical operators more frequently (De Visscher & Noël, 2014; De Visscher et al., 2015). Given these deficits, children with mathematical difficulties find it harder to overcome counting strategies and thus display more error prone processing and have a higher cognitive load in WM resources (e.g., counting strategy with or without fingers). They use these strategies longer than typically achieving children (Geary, 2011).

Arithmetic Production and Verification Tasks

In order to measure mathematical skills in terms of arithmetic fact fluency, there are basically two types of possible tasks: (1) production tasks (e.g., presenting a mathematical problem with limited time for solving the task) and (2) arithmetic verification tasks (e.g., “2 +

2 = 4 : correct?"; Dewi et al., 2021). In contrast to paper and pencil tests, computer-based measurement can provide exact response latencies on the item-level. Practical aspects can limit the validity, though. The computer-based measurement of processing time can have shortcomings if children have to search for the correct digits on the keyboard or screen for typing the solution. The abilities needed in this process, such as familiarity with the keyboard or motor abilities, differ between children in elementary school, which means that processes irrelevant for arithmetic competencies play a role, too, and affect the time and effort needed to carry out the task (Horkay et al., 2006). In contrast, in arithmetic verification tasks, true or false equations are presented that just have to be classified as "true" or "false" by simply tapping one of two buttons on the screen. As a result, we assume that the most valid and robust way to assess arithmetic fact fluency is by means of computer based arithmetic verification tasks.

Arithmetic production and verification tasks differ slightly in their underlying cognitive mechanisms. For arithmetic production tasks, three processing stages may be assumed: (1) encoding of the problem, (2) searching for the answer in long-term memory or solving the task and (3) providing the answer (Ashcraft & Battaglia, 1978). In verification tasks, one additional stage is needed, namely, the evaluation of the presented solution (Ashcraft, 1982; Ashcraft et al., 1984).

Several conditions are known to affect the time needed to solve verification tasks. First, solution times are shorter for true compared to false equations (e.g., Ashcraft & Fierman, 1982; J. I. D. Campbell, 1987). Moreover, subtraction tasks are more difficult to solve than addition tasks, leading to longer solution times (J. I. D. Campbell, 2008; Schneider & Anderson, 2010). Besides that, counting or retrieval processes can be bypassed by using plausibility judgments, which leads to shorter solution times (Reder, 1982). In plausibility judgments, the equation is processed as a whole without performing exact calculations (Zbrodoff & Logan, 1990). Plausibility judgments are more likely (1) if the presented answer differs extremely from the correct answer (Ashcraft & Battaglia, 1978; de Rammelaere et al., 1999; Zbrodoff & Logan, 1990) or (2) if different parities are observed between given and expected answer, such as in 2 + 4 = 7 (Krueger, 1986; Krueger & Hallford, 1984; Lemaire & Fayol, 1995; Lemaire & Reder, 1999; Masse & Lemaire, 2001). Finally, the presented solution in true equations sometimes facilitates the retrieval of the answer (R. N. Campbell, 1978).

In sum, arithmetic verification tasks must be carefully designed to allow meaningful insights into cognitive processes, especially when processing times are used as indicators of arithmetic fluency. However, if successful, arithmetic verification tasks provide an informative and economical possibility for assessing arithmetic fact knowledge in elementary school. Ashcraft et al. (1984) showed that arithmetic production and verification tasks yield converging assessments from Grade 2 on and measure the same skills and constructs.

Research Rationale and Hypotheses

Our aim in this study was to explore the potential of arithmetic verification tasks in elementary school, with a focus of analyzing differences in the processing of arithmetic facts between children with mathematical difficulties and typically performing children. Arithmetic verification tasks have mainly been investigated by using addition and multiplication tasks (e.g., Busch et al., 2013; Busch et al., 2018; Lemaire & Reder, 1999; Widaman et al., 1992; Zbrodoff & Logan, 1990). Schneider and Andersen (2010) used subtraction verification tasks, however, they only investigated adults and only two-digit numbers. In general, less research has been conducted on processing differences between children with mathematical difficulties and typically performing children in elementary school, compared to research addressing children with reading or writing difficulties, and the available research is often limited to single grades (e.g., Busch et al., 2013). Addressing this research gap seems especially important as mathematical difficulties manifest themselves in elementary school. Valid and economic diagnostic tools are needed for identifying children with mathematical difficulties early and providing them with the necessary support (Chodura et al., 2015).

First, to establish construct validity of arithmetic verification tasks, we examined to what extent performance in our newly constructed arithmetic verification task (Richter et al., 2018) corresponded to performance in a standardized arithmetic production test (Study 1). We expected a strong linear relationship between performance in arithmetic production, based on standardized tests and arithmetic verification task in elementary school children (Hypothesis 1). Moreover, we examined the accuracy on the classification of children with mathematical difficulties (*T* score ≤ 40 in standardized tests) by their performance in the arithmetic verification task. To this end, we explored predictive values by ROC analyses (receiver operating characteristic).

Based on the distinction of strategy usage and fact retrieval, we also focused on determinants of accuracy and processing speed in the arithmetic verification task (Study 2). Children with mathematical difficulties should display lower accuracy and should need more time to solve arithmetic tasks (Hypothesis 2). The higher cognitive load induced by task-switching should further impede processing, leading to a decrease of accuracy and an increase in processing time (Hypothesis 3). Since subtraction is acquired later from a developmental perspective, subtraction (vs. addition) items should be more difficult throughout Grades 1 to 4, leading to a decrease of accuracy and an increase in processing time (Hypothesis 4). Moreover, we assumed that children with mathematical difficulties should display a stronger decline of performance (lower accuracy and more time needed to complete the tasks) for two-digit (vs. one-digit) operations than typically achieving children at the end of elementary school (Hypothesis 5). Finally, costs in the task-switching-condition should be higher for children with mathematical difficulties (= lower accuracy and longer response latencies for subtraction vs. addition items; Hypothesis 6).

Study 1

The main objective of Study 1 was to explore the validity of the newly constructed arithmetic verification task. In a controlled setting in elementary schools, we examined (1) to what extent performance in the arithmetic verification task corresponded to performance in a standardized arithmetic production test and (2) how reliably children with mathematical difficulties could be identified by the arithmetic verification task.

Method

Participants

The sample consisted of 165 students recruited from three elementary schools (Grades 2 to 4) in Bavaria, Germany. Gender was balanced in Grade 3 ($n = 50$; 50.0% female), female participants slightly outweighed male participants in Grade 2 ($n = 61$; 57.4% female) and male participants slightly outweighed female participants in Grade 4 ($n = 54$; 44.4% female). Due to a fully anonymized data collection, no additional socio-economic data can be reported. In the participating schools, the age range of children varied between 7-8 years in Grade 2, 8-9 years in Grade 3, and 9-10 years in Grade 4. Data collection took place in a period of two weeks at the end of the school year in July 2019.

Procedure, Design and Instruments

Children were tested together in classrooms of the participating schools. The first measure of arithmetic skills was a newly constructed computerized arithmetic verification task. It was presented on a computer tablet (10.1 inch). First, two instruction items were presented visually and with corresponding audio reading the problem aloud to students. Then, a total of 180 arithmetic items were visually presented in nine units in ascending difficulty. Units 1 to 3 included tasks within number set 1-10 (e.g., "3 + 5 = 8": correct?). Units 4 to 6 tasks spanned number set 1-20 (e.g., "12 - 5 = 7": correct?). The final sets, Units 7 to 9, included tasks within number set -100 (e.g., "36 - 7 = 29": correct?). For the three units in each number set, the first unit only contained addition tasks, the second unit only subtraction tasks and the third unit both addition and subtraction tasks. The students' task was to decide for every single item whether the presented equation was true or false by tapping on a specific area on the tablet (e.g., “2 + 4 = 6” – TRUE; “2 + 4 = 9” – FALSE). Half of the 20 items in each unit were correct. Between every unit there was a short break. The entire administration of the test was limited to 11 minutes. Because of this time limit, only 22.2% of the children in Grade 4 completed the whole task (12.0% in Grade 3; 3.3% in Grade 2) but all children in all grades completed Units 1 to 3 (number set 1-10) and a majority of 88.9% in Grade 4 completed Unit 6 (number set 1-20). Response accuracy and response latencies from presentation onset to tipping one of the response buttons were recorded. The sum of the correctly solved items within 11 minutes served as raw score. Internal consistency (Cronbach’s α) for the whole arithmetic verification task was $\alpha = .98$ (Grade 2: $\alpha = .97$, Grade 3: $\alpha = 0.96$, Grade 4: $\alpha = 0.97$).

To assess arithmetic skills with a standardized arithmetic test, we administered a commonly used mathematics test for elementary school (Heidelberger Rechentest; HRT1-4; Haffner et al., 2005). Three subtests were carried out: addition (e.g., “5 + 3 = _”), subtraction (e.g., “5 – 3 = _”) and fill-in-the-blank (e.g., “6 + _ = 7”). Each subtest consists of a set of 40 computation tasks in increasing difficulty within a time limit of 2 minutes. Children were instructed to work on the tasks in the given order and to solve as many tasks as possible within a given time limit. For the sample reported in the manual of the HRT1-4, test-retest reliability ($r \geq .87$) and the criterion validity were good ($r = .72$ between HRT1-4 and DEMAT 4; Gölitz et al., 2006). The test score of the HRT1–4 was the sum of correct answers.

The convergent validity between HRT 1-4 and the screening procedure LONDI amounts to $r = .772$ in Grade 2, $r = .805$ in Grade 3 and $r = .708$ in Grade 3. Sensitivity and specificity are high ($SN = 85.7\%$; $SP = 93.8\%$), pointing to a high validity of the screening instrument in diagnosing arithmetic disorders in elementary school.

Statistical Analysis

Ordinary least squares (one-level) models were estimated to predict the HRT raw score with the arithmetic verification task raw score. To examine the accuracy on the forecast of children with mathematical difficulties by their performance in the arithmetic verification task, we defined children with mathematical difficulties as having a score on the standardized test that was one standard deviation below the mean (HRT 1-4 T score ≤ 40; below the 16^{th} percentile, respectively). Thus, 21 children in our sample were defined as children with mathematical difficulties. Likewise, children scoring below the 16^{th} percentile in the arithmetic verification task score at each grade level were defined as children at risk. Given that norm-referenced scores were not yet available, we used the present sample as reference group. Based on these cut-off values, children were divided into four groups (Table 2): (1) Children with below average performance in the predictor and criterion variable (true positive; 18 children), (2) children with at least average performance in the predictor and criterion variable (true negative; 135 children), (3) children with below average performance in the predictor variable and at least average performance in the criterion variable (false positive; 9 children) and (4) children with at least average performance in the predictor variable and below average performance in the criterion variable (false negative; 3 children). Next, we calculated the sensitivity (percentage of actual positives correctly identified as such), specificity (percentage of actual negatives correctly identified as such), and RIOC index (relative improvement over chance; see Loeber & Dishion, 1983) for Grades 2, 3, and 4 separately.

Moreover, receiver operating characteristic (ROC) analyses were performed in order to estimate the accuracy on the forecast of children with mathematical difficulties by their performance in the arithmetic verification task. With a value area between 0 and 1 under the ROC curve (with 0.5 as the worst possible value indicating random classification), values near 1 would indicate perfect prediction.

Availability of Data and Materials

All data and analysis scripts are available at the repository of the Open Science Framework (https://osf.io/f8pzk/?view_only=cd81fff16b9348f5995980d13cc753a6). Materials are available from the authors upon request.

Results

A one-way between subjects ANOVA was conducted to compare the effect of grade-level on performance in HRT 1-4 and the arithmetic verification task. There was a significant effect of grade level on both HRT 1-4 raw score, $F(2, 162) = 71.45$, $p < .001$, $\eta^2 = .47$, and on the arithmetic verification task raw score, $F(2, 162) = 37.19$, $p < .001$, $\eta^2 = .32$. Table 1 provides descriptive statistics by grade level. Differences in mathematical performance were greater between Grade 2 and 3 (production task: $d = 1.35$; verification task: $d = 1.02$) than between Grade 3 and 4 (production task: $d = 0.80$; verification task: $d = 0.48$).

[Table 1 near here]

Prediction of the Standardized HRT Test Score by Arithmetic Verification Task Raw Score

Linear regression models were estimated to predict the HRT raw score with the arithmetic verification task raw score as predictor. In line with Hypothesis 1, the model explained a significant and considerable proportion of variance in the HRT raw scores in Grade 2, $F(1, 59) = 86.80$, $p < .001$, $R^2 = .595$, in Grade 3, $F(1, 48) = 88.14$, $p < .001$, $R^2 = .647$, and in Grade 4, $F(1, 52) = 52.16$, $p < .001$, $R^2 = .501$. These findings substantiate the construct validity of the arithmetic verification across elementary school.

Classification of Children with Mathematical Difficulties

When the raw scores of the arithmetic verification task were used to identify children with low arithmetic abilities (determined by performance in the HRT 1-4: T score ≤ 40), high sensitivity (*SN*) and specificity (*SP*) were obtained for all grade-levels: Grade 2 (*SN* = 87.5%; *SP* = 94.3%), Grade 3 (*SN* = 85.7%; *SP* = 95.3%) and Grade 4 (*SN* = 83.3%; *SP* = 91.7%). The *RIOC* indices were high (Grade 2: 85.0%; Grade 3: 83.0%; Grade 4: 80.0%), indicating a highly reliable classification. Moreover, areas under the ROC curves of .91 (Grade 2), .94

(Grade 3) and .92 (Grade 4) were found. Taken together, our arithmetic verification task seems to be an appropriate task for assessing arithmetic abilities and identifying children at risk in elementary school (*SN* = 85.7%; *SP* = 93.8%; *RIOC* = 82.9%; Table 2).

[Table 2 near here]

Study 2

Having established the validity of the newly constructed arithmetic verification task in Study 1, we focused on determinants of accuracy and processing speed in the arithmetic verification task in a larger sample. Given that children can ideally work on tablet-based tasks without supervision, we examined the performance in the arithmetic verification task in an ecologically valid setting: In Study 2, children worked individually at home and the tests were administered online.

Method

Participants

Participants in Study 2 were 3,409 German school children recruited from elementary schools (Grades 1 to 4) in the Federal State of Hesse, Germany. The study was announced through the Ministry of Education of Hesse (a federal state of Germany), which provided elementary schools with the opportunity to participate in a complimentary holiday support program that was accompanied by a screening of reading, writing, and mathematics abilities. The screening included the arithmetic verification task. The elementary schools had to accept participation first and engage their students. Class teachers in participating schools could sign up the children in their class for the screening or only the arithmetic verification task. A total of 772 elementary school classes (out of 11,333 possible classes) participated in our study, with 3.1 to 4.8 children per class on average. In some classes, only individual children were activated for testing by the teachers, while in other classes, it seems that the entire class was activated (up to 20 children). Gender ratio was balanced in each grade level (Grade 1: *n* = 1,034, 50.1% female; Grade 2: *n* = 1,025, 50.5% female; Grade 3: *n* = 999, 50.5% female;

Grade 4: *n* = 351, 48.7% female). For comparing children with sufficient arithmetic skills vs. children with mathematical difficulties, children scoring below the 16th percentile in the arithmetic verification task raw score at each grade level were defined as children with mathematical difficulties, the remaining children represented the control group. We used the arithmetic verification task raw score (sum of the correctly solved items within 11 minutes) as an adequate measure for estimating arithmetic competencies, as indicated by our findings in Study 1. Nevertheless, it is worth noting that the sample size in Study 1 was relatively small (*n* = 165), which should be considered when interpreting the results. We decided to use this cut-off percentile because the 16th percentile refers to a performance 1 *SD* below the mean, which is commonly used as the cut-off percentile representing a performance below average. Consequently, 15.6% of the children were defined as having mathematical difficulties (15.4% in Grade 1, 15.9% in Grades 2 and 3, and 14.5% in Grade 4). Girls were descriptively more often affected than boys in Grade 2 (18.1% vs. 13.6%), in Grade 3 (19.4% vs. 12.3%) and in Grade 4 (16.4% vs. 12.8%), whereas boys were more often affected in Grade 1 (16.9% vs. 13.9%). Since an unexpected gender distribution (more males than females evidenced math difficulties) was observed in the examined sample of Grade 1, we cannot exclude the possibility of selection effects in this sample. Therefore, the findings for Grade 1 should be interpreted with caution.

Design and Instruments

Data collection took place at the end of the school year (July 2020). The tablet-based arithmetic verification task was the same as in Study 1, this time integrated into a screening app for children with learning difficulties (Endlich et al., 2022). The Hessian Ministry of Education and the Arts informed elementary schools in that Federal State of Germany about the possibility to use the screening app and associated trainings for promoting mathematical abilities for free. Thus, the screening app can be downloaded, installed, and freely used at any time in the app store. This offer was made as part of compensatory measures to counteract learning backlogs due to COVID-19 lockdowns in schools during the pandemic. The screening app was intended to represent a low-threshold, voluntary service for all schools. Teachers from elementary schools could encourage their students to download the app and to complete the arithmetic verification task at home or in school on a mobile device. Sample characteristics are provided in Table 3. Given the unproctored nature of the assessment, we are unable to report details regarding the home situation of the children. Additionally, it should be noted that there may have been instances where the task was potentially undertaken by someone other than the intended participant, or where the child may have received external assistance in completing it. Nonetheless, there is reason to believe that such instances were exceptions to the rule and likely occurred infrequently, since the task was presented to the elementary school children by their teacher. This assumption is supported by our data, which shows that the average scores achieved in the samples of Study 1 and Study 2 did not differ substantially from each other (small effect sizes in favor of Sample 2; $d = 0.41$ in Grade 2; $d = 0.22$ in Grade 3; $d = 0.16$ in Grade 4).

[Table 3 near here]

Statistical Analysis and Missing Data

Responses that were unusually slow or fast (3 *SD* or more below the item-specific mean and 2 *SD* or more below or above the person-specific mean after standardizing each item by its item-specific mean) were excluded from the analyses because these responses were likely to be anomalous (comparable to other reaction time studies such as Schindler et al., 2018). Table 4 shows response latencies before and after data exclusion. Given that only very few responses had to be excluded (1,480 or 0.7 % of 204,540 data points), we decided to run the models with data from all participating children, excluding only these unusually slow or fast responses. These exclusions in general did not pose a problem for the analysis, since GLMM is robust against missing data.

[Table 4 near here]

Log-transformed response latencies were analyzed using linear mixed-effects models (LMM: Baayen et al., 2008) with crossed random effects for items nested within participants and participants nested within items, as a considerable amount of variance in the data could be attributed to differences between items and participants (see ICCs in Tables 5 and 6; Baayen et al., 2008). For accuracy data, generalized linear mixed models (GLMM) with a logit link function were estimated, which is the method of choice for nested data structures with binary outcomes (Dixon, 2008). All models were estimated with the software package *lme4* (Bates et al., 2021; Version 1.1-27) for R (Version 4.1.1). For hypothesis tests, we used the software package *lmerTest* (Kuznetsova et al., 2020; Version 3.1-3). All significance tests were based on a Type I error probability of .05. At the beginning of elementary school, children have not yet received instruction to cross the ten barrier and consequently, for the lower grades, a reduced version that included only tasks within the number set up 1–10 was applied. Complete data were available only for number set 1–10 for Grades 1 to 4. Therefore, two separate models were estimated: One for number set 1–10 (Model 1; Grades 1 to 4) and one for number set 1–20 (Model 2; only Grade 4). Intercepts for persons and items were allowed to vary randomly. The following main effects (fixed effects) were included as dummy-coded predictor variables: foundational arithmetic operations (addition = 0, subtraction = 1), switching (standard condition = 0, switch condition = 1) and mathematical difficulties (control group = 0, mathematical difficulties = 1). For Model 1, grade level was centered around 2.5, the mean class level, to model linear developmental trends from Grades 1 to 4. Moreover, interaction effects were estimated for foundational arithmetic operations (addition vs. subtraction) and mathematical difficulties and for switching and mathematical difficulties. The parameter estimates for the fixed and random effects are provided in Table 5 for Model 1 and in Table 6 for Model 2.

Results

Model 1: Number Set 1 to 10

The GLMM for the accuracy data of Units 1 to 3 (Model 1) revealed significant main effects of switching (β = -0.44; z = -2.40; p = .016), mathematical difficulties (β = -0.20; z = -3.28; p = .001) and grade level (β = 0.12; z = 7.06; p < .001). In addition, significant interactions were found for arithmetic operation and mathematical difficulties (β = -0.12; z = -2.16; p = .031) and switching and mathematical difficulties (β = -0.17; z = -2.97; p = .003). The LMM for response latency (Model 1) revealed significant main effects of switching (β = 0.20; t(60) = 5.54; p < .001), mathematical difficulties (β = 0.47; t(3713) = 33.36; p < .001) and grade level (β = -0.20; t(3399) = -39.84; p < .001) and – additionally – arithmetic operation (β = 0.13; t(60) = 3.83; p < .001). Again, significant interactions were found for both arithmetic operation and mathematical difficulties (β = 0.07; t(199769) = 14.59; p < .001) as well as switching and mathematical difficulties (β = 0.07; t(199781) = 12.94; p < .001). Both interactions are depicted in Figure 1: Children with mathematical difficulties spent even more time for items in the task-switching (vs. standard) tasks (Figure 1a) and for subtraction (vs. addition) items (Figure 1b), which supports Hypothesis 6. Whereas typically achieving children spent, on average, 637 ms more time on items in the task-switching (3,361 ms) than standard tasks (2,724 ms), d = 0.33, the difference for children with mathematical difficulties was slightly higher (1,392 ms, d = 0.44): estimated response latencies were 5,884 ms in the task-switching and 4,492 ms in the standard switching condition. The same interaction effect was observed regarding addition vs. subtraction items. Whereas typically achieving children spent, on average, 394 ms more time on subtraction items (3,229 ms) than on addition items (2,836 ms), d = 0.23, children with mathematical difficulties needed 1,035 ms more time for subtraction items (5,710 ms) than for addition items (4,675 ms), d = 0.35.

In line with Hypothesis 2, children with mathematical difficulties displayed lower accuracy and needed more time to solve arithmetic tasks. As predicted by Hypothesis 3, accuracy decreased and processing time increased when it came to the task-switching condition. Moreover, subtraction (vs. addition) items were more difficult throughout Grades 1 to 4 (decrease of accuracy and increase in processing time; Hypothesis 4).

Model 2: Number Set 1 to 20

The GLMM for response accuracy and the LMM for response latency (Model 2) included the ten crossing as additional predictor. As described above, it was based on only the data of Grade 4, representing the end of elementary school. In addition to the predictor single vs. multiple digits, we analyzed the effects of mathematical operation, switching and mathematical difficulties, both for accuracy and response latency as outcome variables. Once again, the interaction between ability level and the other factors was of particular interest, as it represents the surplus in cognitive load in children with mathematical difficulties.

The analysis for response accuracy revealed significant main effects of switching (β = -0.48; z = -3.30; p < .001) and number of digits (β = -0.45; z = -3.25; p = .001). The main effect of arithmetic operation slightly missed level of significance (β = -0.23; z = -1.68; p = .092). The LMM for response latency revealed significant main effects of arithmetic operation (β = 0.16; t(120) = 5.61; p < .001), switching (β = 0.13; t(120) = 4.06; p < .001), mathematical difficulties (β = 0.30; t(385) = 7.07; p < .001) and number of digits (β = 0.25; t(120) = 8.73; p < .001). Moreover, significant interactions were found for mathematical difficulties and arithmetic operation (β = 0.09; t(40589) = 7.07; p < .001), for mathematical difficulties and switching (β = 0.07; t(40589) = 5.20; p < .001) and for mathematical difficulties and number of digits (β = 0.16; t(40685) = 12.34; p < .001), thus supporting Hypothesis 5: Children with mathematical difficulties displayed a stronger decline of performance (lower accuracy and more time needed to complete the tasks) for two-digit (vs. one-digit) operations than typically achieving children at the end of elementary school.

Discussion

The present studies pursued two main goals. First, we examined to what extent performance in our newly constructed arithmetic verification task corresponded to performance in a standardized arithmetic production test (Study 1). Second, we focused on differences between children with mathematical difficulties and typically achieving children with regard to the impact of item-specific characteristics on performance, namely task-switching (switching between arithmetic operations vs. consistent operations), type of arithmetic operation and number set. We were particularly interested in the interaction between difficulty generating factors like multi-digit operations with the aptitude of the children.

Validity of the Newly Constructed Arithmetic Verification Task (Study 1)

We observed large differences in mathematical achievement between different grade levels, in production tasks as well as in verification tasks. Considering the cross-sectional study design, our data cannot provide information about individual developmental trajectories. Nevertheless, given that the development of performance was descriptively comparable for the arithmetic production and the arithmetic verification task (e.g., greater improvement between Grade 2 and 3 than between Grade 3 and 4 in the present study), our results may be regarded as the first evidence for the validity of the arithmetic verification task as a developmentally sensitive measure of mathematical skills. In line with Hypothesis 1 and in accordance with results obtained by Ashcraft et al. (1984), performance in the arithmetic verification task was closely related to performance in an established arithmetic production task and explained between 50% and 65% of the variance in this task within Grades 2 to 4. Moreover, children with mathematical difficulties, identified by below-average scores in arithmetic production tasks (HRT T score ≤ 40), could be reliably identified by their performance in the arithmetic verification task (areas under the ROC curve > .90; RIOC = 82.9%). In sum, the arithmetic verification task raw score can be assumed to be an appropriate estimator for arithmetic performance in elementary school children.

Although arithmetic production and verification tasks differ slightly in their underlying cognitive mechanisms–namely the additional stage in verification tasks, the evaluation of the presented solution (Ashcraft, 1982; Ashcraft et al., 1984)–the basic stages of processing are very similar: (1) encoding of the problem, (2) searching for the answer in long-term memory or solving the task and (3) providing the answer (Ashcraft & Battaglia, 1978). The findings of Study 1 concerning the validity of the newly constructed arithmetic verification task support the idea that arithmetic production and verification tasks measure the same skills and constructs. The results of an additionally conducted confirmatory factor analysis also point to this one-dimensionality (see the Online Supplement at https://osf.io/f8pzk/?view_only=cd81fff16b9348f5995980d13cc753a6).

These findings are of practical importance for the assessment of mathematical skills, as arithmetic verification tasks are far easier to implement in a computer-based fashion than arithmetic production tasks and provide many advantages. Among other things, abilities unrelated to mathematical ability, such as typing skills, play only a minor role in verification tasks, such tasks can be scored automatically and economically and provide not only accuracy data but also precise estimates of response latencies, which may be used as an indicator of processing load.

Performance and Processing Differences Between Children with Mathematical Difficulties and Typically Achieving Children (Study 2)

The results of the linear mixed-effects models (for response latencies) and generalized linear mixed models (for accuracy data) in Study 2 targeted the impact of complexity generating factors on accuracy and time consume and particularly their interaction with the aptitude of children. The underlying assumptions imply an excess in workload for children with mathematical difficulties, indicated by an interaction of item factors with person ability. Again, the analysis revealed the expected substantial increase in children's response accuracy from Grades 1 to 4 and a decrease of the time needed to solve the tasks with increasing grade. Older children may not only use counting strategies more accurately and faster (Widaman et al., 1992), but arithmetic fact knowledge is more and more accumulated over the course of elementary school. Older children can rely on fact retrieval more often and more reliably (Carpenter & Moser, 1984), which at the same time relieves working memory and frees capacities for more complex computations.

Overall, children with mathematical difficulties worked less accurately and spent more time on the tasks (Hypothesis 2). As predicted, task switching interfered more with accuracy and speed in children with problems in mathematics than in typically developing children. The task-switching condition introduced cognitive demands that required individuals to switch between different types of arithmetic operations. This placed additional demands on working memory and cognitive flexibility (Busch et al., 2013). Children with mathematical difficulties may have experienced difficulties in effectively managing these cognitive processes, leading to decreased accuracy and increased processing time in the task-switching condition. The children also spent disproportionately more time on subtraction (vs. addition) tasks (Hypothesis 4). In light of findings on working memory deficits in children with mathematical difficulties (e.g., Schuchardt & Mähler, 2010), this fact suggests a cognitive overload that limits mathematical tasks, even if they are rather simple calculations. Even the load imposed by simply switching between operations hinders mathematical processes in these children. Scenarios that require children with mathematical difficulties to flexibly switch back and forth, build situational models, and combine facts from different sources, such as in word problems, may therefore present an insurmountable challenge.

At the end of elementary school, performance in children with mathematical difficulties catches up to typically achieving children, but only for accuracy and not time consume. Thus, they still needed more time for solving the tasks. These results show quite clearly that children with mathematical difficulties suffer from retrieval deficits that requires them to invest working resources for problems that can be solved easily by children with a normal level of mathematical skills. For mathematics instruction in the final year of elementary school and even more so in secondary school (which starts in Grade 5 in

Germany), this fact poses a challenge. If children with mathematical difficulties need to spend working memory resources for foundational arithmetic operations such as subtraction or simply adapting to a new task, they again lack these resources for more complex mathematical problem-solving activities that become increasingly important in the secondary school. One possible solution would be to make use of focused and comprehensive interventions that are targeted at intense practicing of basic skills of children with mathematical difficulties. These interventions should be applied before children move on to the cognitively more demanding secondary school curriculum.

Conclusions for Assessing Mathematical Difficulties

For assessing mathematical skills, the results imply that reaction times are particularly valuable for identifying children with mathematical difficulties, especially in Grade 4. Neglecting this information can lead to overlooking affected children. The interactions in particular indicate that a highly valuable source of information remains unused if time on task is not recorded. Since children with mathematical difficulties had to spend excessive time for subtraction (vs. addition) tasks (see J. I. D. Campbell, 2008; Schneider & Anderson, 2010), for tasks with two digits (vs. one digit) operations and in the task-switching (vs. standard) condition (Hypothesis 6), these results are predestined to be used in diagnostics. Tests on diagnosing mathematical difficulties should particularly include measures on response latencies, induced by subtraction tasks, two-digit operations and task switching to systematically improve in the diagnoses of mathematical difficulties. At the same time, these effects must be accounted for when tests for mathematical skills involve reaction times, to avoid biases through context effects or differential item functioning.

Limitations and Directions for Future Research

One obvious limitation of this study is that the validity of the newly constructed arithmetic verification task was only shown for Grades 2 to 4 (Study 1). Thus, the results in Study 2 should be interpreted cautiously, especially for Grade 1. The reported similar development of mathematical achievement in production tasks as well as in our verification task between Grades 2 and 4 suggests that the verification task is also valid in Grade 1. Nevertheless, this conjecture needs to be supported in future studies.

As data was assessed online in Study 2, we obviously cannot report details about the situation at children's homes. For example, we cannot exclude the possibility that some children may have had more support than other children or even that another person (e.g., an older sibling) worked on the tasks. The participation rate and the number of students per class mirror the character of that study as an open field study and a free service to the schools. As performance was comparable between the assessments at home (Study 2) and in controlled settings in school (Study 1), it seems plausible to assume that – overall – children worked reliably in the less controlled setting at home. Another limitation of our study is the lack of information about the assignment of participating children to classes or schools. Due to data protection regulations, we were not able to collect further information. As a result, clustering on these levels could not be included in the models.

Regarding the unexpected gender distribution in Grade 1 – more boys than girls showed mathematical difficulties – we cannot exclude the possibility of selection effects in this sample. However, given the uniformity of results across grades, it seems unlikely that gender affected the central results of this study.

Unfortunately, we did not have the opportunity to measure working memory in the present studies. Future research should measure this important construct as a means to back up the interpretation of the present results in terms of cognitive resources. To address the research question whether children with mathematical difficulties require more time for subtraction tasks because they rely on finger counting or other inefficient strategies, future studies could involve interview or observation studies. Moreover, future studies could consider additional information from the school or the teacher regarding whether or not students evidenced mathematical difficulties.

Implications for Practice

The results indicate that children with mathematical difficulties might benefit from fostering their basic arithmetic skills. Improving these skills could reduce the cognitive load and release cognitive resources for demands such as task switching or more complex arithmetic tasks (Zhu & Zhao, 2023).

In terms of practical application, the screening procedure LONDI presented in the article enables an economical testing of an entire school class regarding potential mathematical difficulties. The tablet-based testing can be conducted within half an hour, with immediate and automated evaluation. The screening is already available (Endlich et al., 2022) and is currently undergoing further development.

Conclusion

Our results are relevant both from the perspective of basic research on mathematical difficulties as well as applied psychometrics of the assessment of mathematic skills. With our, from a technical point of view, very simple diagnostic approach, we were able to gather valuable information on the cognitive processes involved in solving arithmetic tasks. We showed differential effects of task switching, arithmetic operations and multi-digit calculations in children with mathematical difficulties compared to normally developing children. We assume that these effects are directly linked to underlying deficits in the routinization of arithmetic procedures and the less efficient access to numerical long term factual knowledge. In that sense, they mirror one of the core problems in children with mathematical difficulties. Thus, process data have a high diagnostic benefit and can further increase the informative value in the construction of test procedures. To our knowledge, these differential effects are not yet widely used to identify children in need for compensatory measures. At the same time, these process data represent a great potential to advance diagnostics, which usually rely only on sum scores of correctly solved items. Future diagnostic procedure could tap on the wealth of process data and especially reaction times and reaction time differences between conditions available in computer-based testing, to further increase the quality of the assessment, beyond gross measures of performance.

References

Amalric, M. & Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage 189*, 19–31. https://doi.org/10.1016/j.neuroimage.2019.01.001

Andersson, U., & Lyxell, B. (2007). Working memory deficit in children with mathematic difficulties: A general or specific deficit? *Journal of Experimental Child Psychology*, *96*(3), 197–228. https://doi.org/10.1016/j.jecp.2006.10.001

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review*, *2*(3), 213–236. https://doi.org/10.1016/0273-2297(82)90012-0

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), 527–538. https://doi.org/10.1037/0278-7393.4.5.527

Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology*, *33*(2), 216–234. https://doi.org/10.1016/0022-0965(82)90017-0

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Bagnoud, J., Dewi, J., Castel, C., Mathieu, R., & Thevenot, C. (2021). Developmental changes on size effects for simple tie and non-tie addition problems in 6- to 12-year-old children and adults. *Journal of Experimental Child Psychology*, *201*. https://doi.org/10.1016/j.jecp.2020.104987

Baroody, A. J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review*, *3*(2), 225–230. https://doi.org/10.1016/0273-2297(83)90031-X

Baroody, A. J. (1984). A reexamination of mental arithmetic models and data: A reply to Ashcraft. *Developmental Review*, *4*(2), 148–156. https://doi.org/10.1016/0273-2297(84)90004-2

Baroody, A. J. (1994). An evaluation of evidence supporting fact-retrieval models. *Learning and Individual Differences*, *6*(1), 1–36. https://doi.org/10.1016/1041-6080(94)90013-2

Baroody, A. J., Tiilikainen, S. H., & Tai, Yu-chi (2006). The application and development of an addition goal sketch. *Cognition and Instruction, 24*(1), 123–170. https://doi.org/10.1207/s1532690xci2401_3

Barrouillet, P., & Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition*, *128*(1), 35–44. https://doi.org/10.1016/j.cognition.2013.02.018

Bates, D., Maechler, M., & Dai, B. (2021). *lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-27*. http://CRAN.R-project.org/package=lme4

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273–293. https://doi.org/10.1207/S15326942DN1903_3

Busch, J., Oranu, N., Schmidt, C., & Grube, D. (2013). Rechenschwäche im Grundschulalter: Reduzierte Verfügbarkeit basalen arithmetischen Faktenwissens und Belastung des Arbeitsgedächtnisses bei Drittklässlern [Dyscalculia in elementary school: Reduced availability of basal knowledge of arithmetic facts and working memory load in third graders]. *Lernen und Lernstörungen, 2*(4), 217–227. https://doi.org/10.1024/2235-0977/a000043

Busch, J., Schmidt, C., Studte, S., & Grube, D. (2018). Kognitive Merkmale rechenschwacher Kinder in Abhängigkeit vom Cut-off Kriterium [Cognitive characteristics of children with mathematical difficulties depending on the cut-off criterion]. *Lernen und Lernstörungen, 8*(3), 167–178. https://doi.org/10.1024/2235-0977/a000258

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*(1), 3–18. https://doi.org/10.1111/j.1469-7610.2004.00374.x

Campbell, J. I. D. (1987). Production, verification, and priming of multiplication facts. *Memory & Cognition*, *15*(4), 349–364. https://doi.org/10.3758/BF03197037

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., & 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

Campbell, J. I. D., & Tarling, D. P. M. (1996). Retrieval processes in arithmetic production and verification. *Memory & Cognition*, *24*(2), 156–172. https://doi.org/10.3758/BF03200878

Campbell, R. N. (1978). *Recent advances in the psychology of language*. Plenum Press

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36*(1), 113-149. http://dx.doi.org/10.1111/j.1469-7610.1995.tb01657.x

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, *15*(3), 179–202. https://doi.org/10.2307/748348

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535–544. https://doi.org/10.1016/S0346-251X(97)00046-8

Chen, Y., & Campbell, J. I. D. (2018). "Compacted" procedures for adults' simple addition: A review and critique of the evidence. *Psychonomic Bulletin & Review*, *25*(2), 739–753. https://doi.org/10.3758/s13423-017-1328-2

Chodura, S., Kuhn, J.-T., & Holling, H. (2015). Interventions for children with mathematical difficulties: A meta-analysis. *Zeitschrift für Psychologie, 223*(2), 129–144. https://doi.org/10.1027/2151-2604/a000211

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of Information Processing* (pp. 151-216). Academic Press.

Crollen, V., & Noël, M. P. (2015). The role of fingers in the development of counting and arithmetic skills. *Acta Psychologica*, *156*, 37–44. https://doi.org/10.1016/j.actpsy.2015.01.007

D'Amico, A., & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences*, *15*(3), 189–202. https://doi.org/10.1016/j.lindif.2005.01.002

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

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *33*(2), 219–250. https://doi.org/10.1016/S0010-9452(08)70002-9

Devine, A., Soltész, F., Nobes, A., Goswami, U., & Szűcs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learning and Instruction, 27*, 31–39. https://doi.org/10.1016/j.learninstruc.2013.02.004

de Visscher, A., & Noël, M.-P. (2014). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal of Experimental Psychology: General*, *143*(6), 2380–2400. https://doi.org/10.1037/xge0000029 de Visscher, A., Szmalek, A., van der Linden, L., & Noël, M.-P. (2015). Serial-order learning impairment and hypersensitivity-to-interference in dyscalculia. *Cognition*, *144*, 38–48. https://doi.org/10.1016/j.cognition.2015.07.007

de Weerdt, F., Desoete, A., & Roeyers, H. (2013). Behavioral inhibition in children with learning disabilities. *Research in Developmental Disabilities*, *34*(6), 1998–2007. https://doi.org/10.1016/j.ridd.2013.02.020

Dewi, J. D. M., 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

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*(4), 447-456. https://doi.org/10.1016/j.jml.2007.11.004

Endlich, D., Lenhard, W., Marx, P., & Richter, T. (2022). *LONDI-Screening: Früherkennung von Problemen im Lesen, Rechtschreiben und Rechnen in der Grundschule* (Beta-Version) [Mobile App]. Meister Cody GmbH. Google Play Store/Apple Store. https://apps.apple.com/de/app/londi-screening/id1517774441 and https://play.google.com/store/apps/details?id=com.meistercody.ferdi

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition*, *123*(3), 392-403. https://doi.org/10.1016/j.cognition.2012.02.008

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

Gaidoschik, M. S. (2010). *Die Entwicklung von Lösungsstrategien zu den additiven Grundaufgaben im Laufe des ersten Schuljahres* [Doctoral dissertation, Universität Wien]. U:theses. https://doi.org/10.25365/thesis.9155

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*(2), 345–362. https://doi.org/10.1037/0033-2909.114.2.345

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental & Behavioral Pediatrics 32*(3), 250–263. https://doi.org/10.1097/DBP.0b013e318209edef

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: a short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology*, *27*(5), 787-797. https://doi.org/10.1037/0012-1649.27.5.787

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging*, *8*(2), 242–256. https://doi.org/10.1037/0882-7974.8.2.242

Gölitz, D., Roick, T., & Hasselhorn, M. (2006). *Deutscher Mathematiktest für vierte Klassen* (DEMAT 4) [German Mathematics Test for Grade 4 (DEMAT 4)]. Hogrefe.

Grube, D. (2006). *Entwicklung des Rechnens im Grundschulalter: Basale Fertigkeiten, Wissensabruf und Arbeitsgedächtniseinflüsse* [Development of calculating in elementary school: Basic skills, knowledge retrieval and influences of working memory]. Waxmann.

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *Heidelberger Rechentest* (HRT 1-4) [Heidelberg calculation test (HRT 4)]. Hogrefe.

Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment, 5*(2). http://www.jtla.org

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research*, *18*(4), 213–221. https://doi.org/10.1111/1540-5826.00076

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, *74*(3), 834–850. https://doi.org/10.1111/1467-8624.00571

Kaye, D. B. (1986). The development of mathematical cognition. *Cognitive Development*, *1*(2), 157–170. https://doi.org/10.1016/S0885-2014(86)80017-X

Koponen, T. K., Sorvo, R., Dowker, A., Räikkönen, E., Viholainen, H., Aro, M., & Aro, T. (2018). Does multi-component strategy training improve calculation fluency among poor performing elementary school children? *Frontiers in Psychology*: *9*(1187). https://doi.org/10.3389/fpsyg.2018.01187

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*(4), 516-531. https://doi.org/10.1016/j.jecp.2009.03.009

Krueger, L. E. (1986). Why 2 x 2 = 5 looks so wrong: On the odd-even rule in product verification. *Memory & Cognition*, *14*, 141–149. https://doi.org/10.3758/BF03198374

Krueger, L. E., & Hallford, E.W. (1984). Why 2 + 2 = 5 looks so wrong: On the odd-even rule in sum verification. *Memory & Cognition*, *12*, 171–180. https://doi.org/10.3758/BF03198431

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2020). *lmerTest: Tests for random and fixed effects for Linear Mixed Effect Models (Lmer objects of lme4 package). R Package version 3.1-2.* http://CRAN.R-project.org/package=lmerTest

Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning and Individual Differences*, *23*, 123–130. https://doi.org/10.1016/j.lindif.2012.08.001

Lemaire, P., & Fayol, M. (1995). When plausibility judgments supersede fact retrieval: The example of the odd-even effect on product verification. *Memory & Cognition*, *23*, 34–48. https://doi.org/10.3758/BF03210555

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, *27*(2), 364–382. https://doi.org/10.3758/BF03211420

Loeber, R., & Dishion, T. J. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*(1), 68–98. https://doi.org/10.1037/0033-2909.94.1.68

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Mabbott, D. J., & Bisanz, J. (2003). Developmental change and individual differences in children's multiplication. *Child Development, 74*(4), 1091–1107. https://doi.org/10.1111/1467-8624.00594

Masse, C., & Lemaire, P. (2001). Do people combine the parity- and five-rule checking strategies in product verification? *Psychological Research*, *65*, 28–33. https://doi.org/10.1007/s004260000030

Mathieu, R., Gourjon, A., Couderc, A., Thevenot, C., & Prado, J. (2016). Running the number line: Rapid shifts of attention in single-digit arithmetic. *Cognition*, *146*, 229–239. https://doi.org/10.1016/j.cognition.2015.10.002

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of*

Educational Psychology. *82*(1), 136–156. https://doi.org/10.1111/j.2044-8279.2011.02033.x

Nunes, T., Dorneles, B. V., Lin, P.-J., & Rathgeb-Schnierer, E. (2016). *Teaching and learning about whole numbers in primary school*. Springer International Publishing. https://doi.org/10.1007/978-3-319-45113-8

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* National Research and Development Center for Adult Literacy and Numeracy.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychological Research*, *62*(1), 72–77. https://doi.org/10.1007/s004260050041

Rapin, I. (2016). Dyscalculia and the calculating brain. *Pediatric Neurology, 61*, 11–20. https://doi.org/10.1016/j.pediatrneurol.2016.02.007

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *89*(3), 250–280. https://doi.org/10.1037/0033-295X.89.3.250

Richter, T., Lenhard, W., Marx, P., & Endlich, D. (2018). Konzeption eines Online-Screenings für Lernstörungen [Outline for an online-screening for learning disorders]. *Lernen und Lernstörungen, 7*(4), 203–207. https://doi.org/10.1024/2235-0977/a000237

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science*, *24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268

Robinson, K. M. (2017). The understanding of additive and multiplicative arithmetic concepts. In D. Geary, D. Berch, R. Ochsendorf, & K. M. Koepke, (Eds.), *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 21–46). Academic Press. https://doi.org/10.1016/B978-0-12-805086-6.00002-3

Robinson, K. M., & Dubé, A. K. (2009). Children's understanding of addition and subtraction concepts. *Journal of Experimental Child Psychology*, *103*(4), 532–545. https://doi.org/10.1016/j.jecp.2008.12.002

Robinson, K. M., & Dubé, A. K. (2012). Children's use of arithmetic shortcuts: The role of attitudes in strategy choice. *Child Development Research*, *2012*, 1–10. https://doi.org/10.1155/2012/459385

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1–36.

Rourke, B. P. (1993). Arithmetic disabilities, specific and otherwise: A neuropsychological perspective. *Journal of Learning Disabilities, 26*(4), 214–226. https://doi.org/10.1177/002221949302600402

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27*(4), 763–797. https://doi.org/10.1037/0096-1523.27.4.763

Schneider, D. W., & Anderson, J. R. (2010). Asymmetric switch costs as sequential difficulty effects. *Quarterly Journal of Experimental Psychology*, *63*(10), 1873-1894. https://doi.org/10.1080/17470211003624010

Schindler, J., Richter, T., Isberner, M-B., Naumann, J., & Neeb, Y. (2018). Construct validity of a process-oriented test assessing syntactic skills in German primary schoolchildren. *Language Assessment Quarterly, 15*(2), 183–203. https://doi.org/10.1080/15434303.2018.1446142

Schuchardt, K., & Mähler, C. (2010). Unterscheiden sich Subgruppen rechengestörter Kinder in ihrer Arbeitsgedächtniskapazität, im basalen arithmetischen Faktenwissen und in den numerischen Basiskompetenzen? [Do subgroups of children with arithmetic impairment differ in their working memory capacity, basal knowledge of arithmetic facts, and basic numerical skills?] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *42*(4). https://doi.org/10.1026/0049-8637/a000022

Shalev, R. S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology, 47*(2), 121–125. https://doi.org/10.1111/j.1469-8749.2005.tb01100.x

Siegler, R. S. (1991). C*hildren's thinking*. Prentice-Hall.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press. https://doi.org/10.1093/oso/9780195077872.001.0001

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Erlbaum Associates.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–294). Erlbaum.

Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., & Henschel, S. (2022). *Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe: Erste Ergebnisse nach über einem Jahr Schulbetrieb unter Pandemiebedingungen* [Skills in the subjects German and Mathematics at the end of Grade 4: First results after more than a year of schooling under pandemic conditions]. Institut zur Qualitätsentwicklung im Bildungswesen.

Thevenot, C., Barrouillet, P., Castel, C., & Uittenhove, K. (2016). Ten-year-old children strategies in mental addition: A counting model account. *Cognition*, *146*, 48–57. https://doi.org/10.1016/j.cognition.2015.09.003

Uittenhove, K., Thevenot, C., & Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? *Cognition*, *146*, 289–303. https://doi.org/10.1016/j.cognition.2015.10.008

UNESCO Institute for Statistics. (2017). *More than one-half of children and adolescents are not learning worldwide*. Fact Sheet No. 46, September 2017. UNESCO Institute for Statistics. Retrieved from: http://uis.unesco.org/sites/default/files/documents/fs46-more-than-halfchildren-not-learning-en-2017.pdf

van der Sluis, S., De Jong, P. F., & van der Leij, A. V. D. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology*, *87*(3), 239-266. https://doi.org/10.1016/j.jecp.2003.12.002

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

Widaman, K. F., Little, T. D., Geary, D. C., & Cormier, P. (1992). Individual differences in the development of skill in mental addition: Internal and external validation of chronometric models. *Learning and Individual Differences*, *4*(3), 167–213. https://doi.org/10.1016/1041-6080(92)90002-V

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities*, *34*(3), 237-248. https://doi.org/10.1177/002221940103400304

Zaunmüller, L., Domahs, F., Dressel, K., Lonnemann, J., Klein, E., Ischebeck, A., & Wilmes, K. (2009). Rehabilitation of arithmetic fact retrieval via extensive practice: A combined fMRI and behavioural case-study. *Neuropsychological Rehabilitation*, *19*(3), 422–443. https://doi.org/10.1080/09602010802296378

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), 83-97. https://doi.org/10.1037/0278-7393.16.1.83

Zhu, X. & Zhao, X. (2023). Role of executive function in mathematical ability of children in different grades. *Acta Psychologica Sinica, 55*(5), 696–710. https://doi.org/10.3724/SP.J.1041.2023.00696

Table 1

Descriptive Statistics and Correlation Coefficients for Production (HRT 1-4) and Arithmetic Verification Tasks at Grade 2, 3 and 4 (Study 1)

		N	*M*	*SD*	*r*
Grade 2	HRT *T*-value	61	50.05	9.71	
	HRT (raw score)	61	46.62	13.13	
	AVT	61	101.21	24.80	.772***
Grade 3	HRT *T*-value	50	50.16	10.32	
	HRT (raw score)	50	64.38	13.12	
	AVT	50	127.08	26.01	.805***
Grade 4	HRT *T*-value	54	50.87	9.86	
	HRT (raw score)	54	74.24	11.41	
	AVT	54	140.28	23.65	.708***
Overall					.843***

Note. The reported data represents the raw data of both tests. Correlations calculated between raw scores of HRT 1-4 and the arithmetic verification task (AVT).

*** $p < .001$.

Table 2

Arithmetic Verification Task as Predictor for Children with Mathematical Difficulties Assessed with HRT 1-4 (Study 1)

		Criterion variable: arithmetic production tasks (HRT 1-4)	
		Children with mathematical difficulties (*T*-score < 40)	Children with sufficient arithmetic skills (*T*-score ≥ 40)
Predictor variable: arithmetic verification task	< 16th percentile	18[a]	9[b]
	≥ 16th percentile	3[c]	135[d]
Predictive values	Sensitivity	85.7%	
	Specificity	93.8%	
	Positive predictive value	66.7%	
	RIOC	82.9%	

Note. *N* = 165. HRT 1-4 = Heidelberger Rechentest (Haffner et al., 2005); RIOC = relative improvement over chance;

[a]true positive; [b]false positive; [c]false negative; [d]true negative.

Table 3

Sample Characteristics for Study 2

	Participating classes	Number (*M* and *SD*) of participating children per class	Range of participating children per class (min–max)	Gender (% female)
Grade 1	220	4.70 (4.00)	1–17	50.10
Grade 2	212	4.83 (4.26)	1–20	50.54
Grade 3	226	4.42 (3.60)	1–18	50.45
Grade 4	114	3.08 (2.88)	1–13	48.72

Table 4

Descriptive Statistics for Response Accuracy and Response Latency (Raw Score in ms and Log-Transformed) as Dependent Variables in the Arithmetic Verification Task

	Response accuracy[a]		Response latency (ms)		Response latency (log-transformed)		Number of observations
	M	*SD*	*M*	*SD*	*M*	*SD*	
Number set 1–10							
Before data exclusion	0.95	0.21	4676	327935	8.03	0.61	208,010
After data exclusion	0.95	0.21	3768	3283	8.04	0.58	203,060
Number set 1–20							
Before data exclusion	0.94	0.24	5403	210987	8.21	0.70	87,498
After data exclusion	0.95	0.22	3979	3538	8.10	0.57	86,836

Note. [a] Proportions.

Table 5

Fixed Effects and Variance Components in the Generalized Linear Mixed Model for Response Accuracy and in the Linear mixed-effects Model for Response Latency for Grades 1 to 4 (Model 1).

Parameter	Responce accuracy	Response latency
	β (*SE*)	β (*SE*)
	Fixed effects	
Intercept	3.761 (0.14)***	7.784 (0.03)***
Arithmetic operation	-0.044 (0.18)	0.133 (0.03)***
Switching	-0.443 (0.18)*	0.204 (0.04)***
Mathematical difficulties	-0.203 (0.06)**	0.466 (0.01)***
Arithmetic operation X mathematical difficulties	-0.120 (0.06)*	0.069 (0.00)***
Switching X mathematical difficulties	-0.167 (0.06)**	0.066 (0.01)***
Grade level	0.123 (0.02)***	-0.201 (0.01)***
	Variance components	
Subjects	0.549	0.330
Items	0.454	0.073

Note. Grade level is centered around 2.5. Operation: dummy-coded (addition = 0, subtraction = 1). Switching: dummy-coded (consistent arithmetical operations; only addition OR subtraction tasks in one unit = 0, switching between operations within one unit = 1). Mathematical difficulties: dummy-coded (control group, percentile ≥ 16 in arithmetic verification task = 0; children at risk, percentile < 16 in arithmetic verification task = 1).

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed). Number of observations (participants X items) = 203,060.

Table 6

Fixed Effects and Variance Components in the Linear Mixed-Effects Model for Response Latency and in the Generalized Linear Mixed Model for Response Accuracy for Number Set 1–20 at the End of Elementary School (Model 2).

	Responce accuracy	Response latency
Parameter	β (*SE*)	β (*SE*)
	Fixed effects	
Intercept	4.058 (0.14)***	7.567 (0.03)***
Arithmetic operation	-0.231 (0.14)	0.163 (0.03)***
Switching	-0.476 (0.14)***	0.125 (0.03)***
Number of digits	-0.446 (0.14)**	0.253 (0.03)***
Mathematical difficulties	-0.131 (0.19)	0.298 (0.04)***
Operation X mathematical difficulties	0.021 (0.14)	0.086 (0.01)***
Switching X mathematical difficulties	-0.210 (0.15)	0.071 (0.01)***
Number of digits X mathematical difficulties	-0.163 (0.15)	0.158 (0.01)***
	Variance components	
Subjects	0.610	0.281
Items	0.465	0.096

Note. Operation: dummy-coded (addition = 0, subtraction = 1). Switching: dummy-coded (consistent arithmetical operations; only addition OR subtraction tasks in one unit = 0, switching between operations within one unit = 1). Mathematical difficulties: dummy-coded (control group, percentile ≥ 16 in arithmetic verification task = 0, children at risk, percentile < 16 in arithmetic verification task = 1).

* *p* < .05, ** *p* < .01, *** *p* < .001 (two-tailed). Number of observations (participants X items) = 19,953.

Figure 1

Interactions Between (a) Switching Conditions and Mathematical Difficulties (MD) and (b) Arithmetic Operations and Mathematical Difficulties (MD) for Response Latencies