

# Assessment of Computation Competence and Non-Count Strategy Use in Addition and Subtraction in Grade 1

Delia Leuenberger<sup>1,2§</sup> , Elisabeth Moser Opitz<sup>2§</sup> , Noemi Gloor<sup>2</sup> 

[1] University of Applied Sciences of Special Needs Education, Zurich, Switzerland. [2] Institute of Education, University of Zurich, Zurich, Switzerland.

§These authors contributed equally to this work.

Journal of Numerical Cognition, 2024, Vol. 10, Article e12633, <https://doi.org/10.5964/jnc.12633>

Received: 2023-08-18 • Accepted: 2024-03-12 • Published (VoR): 2024-04-17

Handling Editor: Andreas Obersteiner, Technical University Munich, Munich, Germany

Corresponding Author: Elisabeth Moser Opitz, University of Zurich, Institute of Education, Freiestrasse 36, 8032 Zurich, Switzerland. E-mail: [elisabeth.moseropitz@uzh.ch](mailto:elisabeth.moseropitz@uzh.ch)

## Abstract

Computation competence (CC) in simple addition and subtraction using non-counting (NC) strategies is an important learning objective in Grade 1 mathematics but many children, especially low achievers in mathematics, struggle to acquire these skills. To provide these students with the support they need, it is important to have valid and reliable tools for assessing progress in CC and NC strategy use. Developing an assessment instrument for use in Grade 1, when some children start the year unable to solve any problems, is challenging, as is ensuring measurement invariance over a school year when children generally make large achievement gains. This paper presents a new assessment tool for CC and NC strategy use in Grade 1 that was tested in a longitudinal study with  $N = 1,017$  children. Analyses using the Rasch model revealed acceptable mean square scores (MNSQ 0.83 – 1.20). Warm's Weighted Likelihood Estimate (WLE) reliability scores were acceptable (pre-test .77; post-test .87). Measurement invariance over time was given. The instrument is promising for assessing CC and NC strategy use efficiently and accurately in Grade 1.

## Keywords

computation competence, assessment, measurement invariance over time, non-counting strategies, persistent counting

## Non-Technical Summary

### Background

A key objective of first grade mathematics is to teach children strategies for solving simple addition and subtraction problems without counting by, for example, being able using decomposition strategies. Unfortunately, some children, especially low achievers, struggle to learn these non-counting strategies.

### Why was this study done?

Researchers and educators need valid and reliable tools to assess student computation proficiency and use of non-counting strategies so that they can evaluate progress over the course of the year and identify students who need extra support.

### What did the researchers do and find?

We developed and tested a diagnostic tool for assessing the computation proficiency and strategy use of first graders over the course of the school year.  $N = 1,017$  children participated in the study. They were asked to solve a series of simple math



problems (e.g.  $6 + 3$ ;  $7 - 3$ ) in their heads while tapping on the table with the palm of their hand or a finger. The way the tapping was performed was useful in identifying counting strategies. For example, an irregularity in the tapping suggested the use of a hidden counting strategy. This could be checked by asking the child, “How did you figure out the problem?”

The results showed that the diagnostic tool is reliable for identifying when children use counting strategies. It was also found that the use of non-counting strategies increased over the course of Grade 1. The greatest increase in use of non-counting strategies was seen for calculations where the power of five, decomposition of tens, or doubling could be used. More difficult problems tended to show a smaller, but still substantial, increase in the use of non-counting strategies.

### What do these findings mean?

The study demonstrated that this diagnostic tool can accurately assess the development of computation proficiency and use of non-counting strategies in Grade 1. It can therefore be used by researchers to investigate how computation proficiency and strategy use develops and by practitioners to identify which first grade students may need extra support to learn how to calculate with non-counting strategies.

### Highlights

- The article presents a reliable test for distinguishing between counting and non-counting strategy use in Grade 1.
- The instrument is suitable for assessing the development of non-counting strategies at the beginning and at the end of Grade 1.
- The frequency with which first-graders use non-counting strategies to solve math problems and the frequency of correct answers are highly correlated.
- The use of non-counting strategies increased the most in the course of Grade 1 for math problems that can be solved using the power of five, decomposition of tens, or doubling.

Learning to add and subtract whole numbers using computation strategies and fact fluency is the basis of successful mathematical learning and an important goal of Grade 1 education (Kilpatrick et al., 2001; Morano et al., 2020; Sievert et al., 2021; Verschaffel et al., 2007). According to Cowan (2003) “a child who cannot efficiently produce the sums of the basic combinations is at a disadvantage in multidigit written and oral arithmetic” (p. 43). But a significant number of children do not succeed in moving on from counting and developing alternative strategies over the course of elementary school. Gaidoschik (2012) found that up to 27% of children continued to favor counting strategies for sums and minuends to 10 at the end of Grade 1; only one third were able to fully automatize operations to 10. Hopkins et al. (2022) reported that third graders used the min-counting strategy (counting on from the larger summand) in 17% of trials when performing basic addition. These children often made mistakes when forced to use a retrieval strategy. A study by Hopkins and Bayliss (2017) found that just over half of seventh graders were still using min-counting to solve simple addition problems, about 30% of whom usually arrived at incorrect answers. The reliance on counting strategies is particularly evident in students with mathematical learning difficulties and low achievers in mathematics who also have difficulty acquiring computational fluency using decomposition or retrieval strategies (Jordan et al., 2003; Kilpatrick et al., 2001; Morano et al., 2020; Moser Opitz, 2013; Verschaffel et al., 2007). Hopkins et al. (2022), Sievert et al. (2021) and Gaidoschik (2012) highlight the importance of supporting the acquisition of efficient NC strategies from the earliest stages of mathematics education. “Reliance on computing by counting beyond the first months of grade one will hamper arithmetic development notably, because it impedes gaining an understanding of both number structures and relationships between operations” (Gaidoschik, 2012, p. 310).

It is important to develop diagnostic tools that can reliably assess student progress in using NC strategies. However, developing an assessment tool is not straightforward. Some children – usually low achievers – cannot solve addition and subtraction problems when they start first grade (Moser Opitz, 2001), making it impossible to establish baseline values. There are also issues that arise due to the requirements of statistical analysis. As far as we are aware, there are only two studies that report reliability scores when investigating strategy use (Hopkins et al., 2022; Wittich, 2017). An instrument

would also have to measure the same construct at both measurement points; have measurement invariance over time (Putnick & Bornstein, 2016). To the best of our knowledge, no such instrument has been developed to date.

This study presents a new diagnostic tool for assessing computation competence (CC) and NC strategy use in Grade 1. CC is defined as correctly solving basic addition and subtraction problems using NC strategies. A sample of  $N = 1,017$  first graders was studied and the relationship between correct solution and NC strategy use was analyzed using the partial-credit Rasch model (Wright & Masters, 1982).

## Computation Competence and Non-Counting Strategies in Addition and Subtraction

### From Counting All to Efficient Computation Strategies

Young children start by spontaneously using informal counting strategies to calculate and then progress from counting-all to counting-on using their fingers or manipulatives. Counting-all and counting-on are distinct from finger display, where children use their stretched fingers statically as a structured set, recognizing the rule of five. Counting strategies are eventually replaced by addition-related calculations (repeated adding, doubling) and derived-fact or decomposition strategies (e.g.,  $6 + 7 \rightarrow 6 + 6 = 12 \rightarrow 12 + 1 = 13$  or  $8 + 6 = 8 + 2 + 4 = 14$ ) based on basic arithmetic principles (Carpenter & Moser, 1984; Clements & Sarama, 2007; Cowan, 2003; Kilpatrick et al., 2001; Verschaffel et al., 2007). Eventually, children are able to recall the result of a computation problem from a retrieval network (De Smedt et al., 2010; Verschaffel et al., 2007). This acquisition of computational fluency enables children to “allocate more cognitive resources to other aspects of the problem, which is particularly important in the use of memory-intensive strategies, such as decomposition” (Vasilyeva et al., 2015, p. 1490). The phases overlap (Baroody et al., 2014) and children often use old counting strategies alongside newer, NC strategies (Shrager & Siegler, 1998). Counting strategies are therefore important in the acquisition of computation competence in young children and also useful for solving addition and subtraction problems using small numbers. But, as outlined in the introduction, a strong reliance on counting strategies can have negative outcomes: It is error prone when computing with higher numbers, and counting hampers pupils’ understanding of the place value system (Gaidoschik, 2012). Unlike in other studies, which use interviews to examine how children choose a strategy or discover new ones (Häsel-Weide, 2016; Siegler & Jenkins, 1989), in this paper we focus on differentiating reliably between counting and NC strategies and the longitudinal assessment of these strategies.

### Assessing Computation Competence and Strategy Use

Studies show that assessing CC without investigating the use of strategy provides little information about the actual arithmetical competence of children (Moser Opitz, 2001; Wittich, 2017). Observing strategy use provides important insights into the development of mathematical understanding. However, identifying strategy use, especially in counting, is difficult because children sometimes count very quickly in their heads or cleverly hide counting strategies, especially when calculating single-digit numbers. Researchers have used a combination of observation and verbal reporting (Gaidoschik, 2012) or a combination of observation, reaction time, and verbal reporting to capture calculation strategies (Hopkins et al., 2022; Siegler, 1987; Vanbinst et al., 2014). However, these studies were, with the exception of Gaidoschik (2012) and Siegler (1987), conducted with students in third through seventh grade. Studying NC strategy use in younger children is more difficult. When asked to describe their strategy, they can struggle to find the right words (Siegler, 1987). Gaidoschik (2012) reported that children answered “I don’t know” when asked how they had solved a problem in 25% of the trials.

Another challenge specific to the assessment of CC (and strategy use) at the beginning of Grade 1 is that at that point the children have had no formal mathematical education. Although most first graders already have significant arithmetic skills when they start formal education (Deutscher, 2012) some are unable to solve any addition or subtraction problems (Moser Opitz, 2001), making it impossible to determine a baseline for strategy use.

Another methodological issue is the reliability of measurements of strategy use. Studies based on the observation of computation strategies seldom calculate reliability scores. Good (satisfying) reliability scores were reported by Hopkins et al. (2022) and Wittich (2017). But Hopkins et al. (2022) examined the strategy use of third and fourth graders and Wittich (2017) conducted the study with second graders. Repeated measurements and calculations of difference values also result in an accrual of measurement errors and a consequent decrease in reliability (Rost, 2004). An instrument

should measure the characteristic of interest independent of other relevant variables, such as time (Putnick & Bornstein, 2016). In other words: “The measurement structure of the latent factor and their survey items should be stable, that is ‘invariant’” (Van de Schoot et al., 2015, p. 1). Differential development at the indicator level must be ruled out to ensure that the same construct is measured across the different measurement points. Achieving measurement invariance over time when assessing CC and strategy use in Grade 1 is complicated by the children having a wide range of abilities. Plus, there is usually a big improvement in mathematics achievement over the course of the first year of school. If an instrument is suitable for assessing competence at the end of Grade 1, there is a high likelihood that it will be too difficult for children at the beginning of that year. If the same items are used at both measurement points, the items must cover a wide range of difficulty. Unfortunately, this increases the risk that the items and the test do not meet psychometric requirements.

All of this suggests that it is important to modify existing methods used to identify and assess counting and NC strategies so that they are reliable when used with young children. Wittich (2017) and Grube and Seitz-Stein (2012) used a dual-task method to detect counting strategies among young children. The method asks children to do a second task (tapping with a finger or the hand on the table, reciting nonsense words such as de-de-de) while calculating. Wittich (2017) combined observation, reaction time, verbal reporting, and tapping to assess calculation strategies in second graders. Results indicated that the way the tapping was performed was useful in identifying that a student was using a counting strategy. For example, when a child started tapping irregularly when calculating or when the tapping process was aborted, this was an indication that a hidden counting strategy was being used. This could be verified by asking the child “How did you figure out the problem?”

Finally, it is important to consider the difficulty and characteristics of the math problems used to assess CC and strategy use (Siegler, 1987). A child’s choice of strategy depends on the characteristics of the math problem. The probability of solving an addition problem with a small second summand (e.g.,  $5 + 2$ ) or doubling 4 with an NC strategy is higher than for  $7 + 8$ , for example, so it is also important to look at the relationship between item characteristics and strategy use over the course of a school year.

## Study Objectives

The literature review revealed that it is difficult to assess achievement gains in CC and increased use of NC strategies over the course of first grade. Wittich (2017) developed an instrument for Grade 2 that assessed CC and NC strategy use in basic addition and subtraction, but the study did not look at measurement invariance over time. To our knowledge, no previous study has considered measurement invariance when assessing the CC and NC strategy use of young children asked to solve basic arithmetic problems. Neither have researchers analyzed the longitudinal relationship between the difficulty level of the item and the use of NC strategies by young students. We have now developed a new tool based on the work of Wittich (2017) that aims to assess CC and NC strategy use at the beginning (t1) and end (t2) of Grade 1 while ensuring reliability, validity, and measurement invariance over time. The following research questions are addressed:

RQ1: How can children’s CC and NC strategy use in basic addition and subtraction be assessed in a valid and reliable way over the course of first grade?

RQ2: How are item difficulty and the use of NC strategies related?

The instrument has been developed and piloted in several steps. In this paper we report the instrument in its final form.

## Method

### Participants

First graders ( $N = 1,017$ ; 49.5% girls,  $M_{\text{age}} = 6.82$ ,  $SD = 0.38$ ) from 69 classes in German-speaking Switzerland participated in an intervention study designed to support the development of NC strategies and social participation (Leuenberger, 2021). All classes used a textbook with a focus on flexible strategy use. Invitation letters were sent to schools and

teachers voluntarily decided whether to participate in the study. Parents had to provide written consent for the study. German was the first language for 79.6% of the children. All of the children were tested at the beginning (t1, August/September) and end (t2, May/June) of the school year. The project had the approval of the ethical committee of the Faculty of Arts and Social Sciences of the University of Zurich.

## Development of the Instrument

### Pre-Test

To avoid frustrating children unable to solve addition and subtraction problems at the beginning of Grade 1, we administered a simple arithmetic pre-test to all participants. A set of correct and incorrect math problems ( $4 + 2 = 6$ ,  $2 + 4 = 6$ ,  $2 + 3 = 5$ ,  $3 + 2 = 6$ ,  $5 + 3 = 9$ ,  $3 + 5 = 8$ ) was presented on the screen and read aloud by the test administrator. For each problem, the child had to say whether the sum was correct or incorrect. To reduce the influence of guessing, the same math problem was presented twice with reversed summands and a different result. If the child solved both problems of a pair with reversed summands correctly, it was scored 1. The child's score was reduced by 25% to adjust for guessing. Children who scored 0 on the pre-test ( $n = 154$ , 15%) scored 0 on the CC test at t1 (they did not sit the test); they participated in the CC test at t2.

### Computation Test

As outlined in the introduction, using the same instrument to assess the numerical competence of children over the course of a school year in Grade 1 is challenging because of the rapid pace of development at that age (Aunio et al., 2015). To address this problem, we used the same test twice, but added three items with a higher difficulty level at t2. The addition and subtraction problems were based on an instrument by Wittich (2017) (Table 1). Several pilot tests were conducted to determine whether the test is suitable for assessing the use of CC and NC strategies in first graders over the course of a school year (Leuenberger, 2021). The final scale for t1 had 17 items.

**Table 1**

*Addition and Subtraction Problems*

Addition	Subtraction
$5 + 2$	$5 - 3$
$6 + 3$	$7 - 3$
$4 + 6$	$10 - 4$
$3 + 15$	$11 - 3$
$7 + 7$	$19 - 4$
$9 + 5$	$14 - 5$
$17 + 4$	$9 - 6$
$6 + 12$	$16 - 8$
$8 + 7$	$23 - 6^a$
$8 + 23^a$	$20 - 13^{ab}$

<sup>a</sup>Items were only used at t2. <sup>b</sup>Complex problem (two numbers with two digits): Answer requested within 5 seconds.

The addition and subtraction problems were presented as 1.5 cm tall black digits separated by spaces on a grey background. The items were presented to children in the sequence in which they are shown in Table 2, according to the relative difficulty of the items in the pilot test. To ensure that the evaluation of counting strategies was as reliable as possible, administrators used observation, solution time, verbal reports, and tapping. The administrator asked the child to solve the problems. They did not give the child any information about how the problems should be solved but did instruct them to tap on the table with the palm of their hand or a finger at a pre-set rate during the computation process (approximately 120 beats per minute; Wittich, 2017).

The coding scheme differentiated between a correct solution (1) and an incorrect one (0). Strategy use was assessed independently of the correctness of the answer and the aim was to differentiate between counting and NC strategies (statically using stretched finger pictures, retrieval, decomposition  $\rightarrow 8 + 7 \rightarrow 8 + 8 - 1$ ). Counting strategies were scored with 0, NC strategies with 1. If the child gave an answer within three seconds (or five for more complex problems, see Table 1) and there was no observable indication of counting, the strategy was scored as NC. When a counting strategy was clearly observable (finger counting, moving the lips, verbal counting), the strategy was scored as 0. The tapping helped observers to recognize hidden counting strategies. If the tapping was irregular or stopped after the problem was presented, it was assumed that the child might have counted. In this case – and also when the strategy could not be identified – the children were asked “How did you figure it out?” A maximum of two points could be scored for each item (1 point for a correct solution, an additional point for an NC strategy; partial credit scoring). If a problem was solved incorrectly using an NC strategy, it was scored 0. If a child gave no answer, the item was also scored 0. In some cases, it was not possible to determine what strategy had been used (3-15% per item at t1, 2-8% per item at t2). If the result was correct in these instances, strategy was scored as Missing (999). If a child was unable to solve three consecutive addition items (incorrect or no answer), the unsolved addition items were scored 0 and subtraction items were presented. If the child was unable to solve three consecutive subtraction items (incorrect or no answer), the test was halted (termination criterion), and the unsolved items were scored 0.

The test was administered by the doctoral students working on the project and other test administrators who were all trained in how to evaluate NC strategy use in intensive role-playing sessions. The trainers used various counting and NC strategies to solve the problems and trainees were asked to identify them. If the trainees did not agree on what strategy was being demonstrated, the assessment decisions were discussed until a consensus was reached. If there was any uncertainty about NC strategy use during the test, the administrator coded it as “unclear”. Notes were made for all observations.

Each child was individually tested in a quiet room in the school building. The child sat to the left of the test administrator so that the administrator could control the program without blocking the child’s view of the screen. The child was read a standardized introduction before each new type of problem and asked to give verbal answers. Data entry (responses and strategies) was exclusively carried out by the test administrator. Test duration was 10 to 15 minutes and the time taken to complete each problem was recorded by the program. At the end of each test session, the program automatically saved the data set for each child.

## Analyses

### Item Difficulty, Reliability, and Measurement Invariance

The data were analyzed using a partial credit Rasch model (Rasch, 1960; Wright & Masters, 1982). The Rasch model provides information about “how well each item fits within the underlying construct” (Bond & Fox, 2001, p. 26). The model estimates the probability of a person with a given ability-level answering a problem correctly. The person’s ability-level and the item’s difficulty are located on the same scale (Bond & Fox, 2001). It is expected that a person with a higher ability level has a greater probability of answering difficult items correctly than a person with a lower ability level. In the dichotomous model, items are scored pass/fail (0/1). In a partial credit model, intermediate levels are scored (Wright & Masters, 1982). An item scored with 0, 1, 2 is handled as a two-step item. The first step is to solve the item correctly and the second step is to solve it using an NC strategy. The partial credit model estimates a parameter for the ability of the person, the difficulty parameter of the item, and thresholds (Schwab & Helm, 2015). Thresholds are boundaries between categories (in this instance, between the scores 0, 1, and 2) and correspond to the point where the probability of solving an item in that category or one above is equal to that of solving it in a category below (Wilson, 2005). The interpretation of the threshold corresponds to the interpretation of the difficulty parameters (Wilson, 2005).

Item fit was measured with a mean square score (MNSQ). Wilson (2005) considers MNSQ values between 0.75 and 1.33 acceptable. To assess measurement invariance over time, DIF analyses were conducted using Masters’ (1982) partial credit model. An incidence of DIF means that the difficulty of certain items varies over time relative to the majority of the items. The analysis measures deviation of the calculated parameter at the item level from the ideal of a zero-difference with time as the main effect. According to Paek and Wilson (2011), a DIF greater than 0.638 logits ( $p <$

.05) between the parameters of the item at t1 and t2, is considered large and a DIF less than 0.638 ( $p < .05$ ) is medium. If no significance is given ( $p \geq .05$ ), differences up to 0.426 can be ignored. It is recommended that items with a double deviation higher than 0.638 be further examined at a given level of significance ( $p < .05$ ) and eliminated from the test if necessary (Paek & Wilson, 2011). Analyses were performed using R statistical software (version 3.6.1) and the TAM package (Robitzsch et al., 2020).

To assess measurement invariance over time the three-dimensional data structure was transformed into a two-dimensional structure. The data matrix for t2 was vertically appended to the matrix for t1 (Hartig & Kühnbach, 2006). By treating the measurement of the same individuals at a later measurement point, virtual individuals were created. Warm's weighted likelihood estimate (WLE; Warm, 1989) person ability parameters were calculated using this newly derived data set. Item difficulties across measurement points were equalized so that the person ability parameters could be placed on the same scale. As a third step and for the further repeated-measures analyses, person abilities parameters were realigned to the original data structure (each person in a row with measurement-time specific ability parameters in columns).

### Validity

Criterion-related validity was evaluated by assessing early numeracy competence as well addition and subtraction competence at the beginning of Grade 1. Because a group test for children at the beginning of first grade was not available, a numeracy test developed by the research team was administered to groups of 10 children who were guided by a trained test administrator (Leuenberger, 2021). The test had 31 items. Twenty-four of the items assessed the topics completing a number sequence for digits up to 20; comparing numbers up to 20 (e.g., 13/14); the part/whole relationship (allocating six sheep to two pastures in different ways); counting objects and matching the correct term to a picture (two red balloons and three blue balloons  $\rightarrow 2 + 3$  or  $3 + 2$ ). The other seven items covered addition and subtraction and included adding coins (e.g., 2 Swiss francs plus 1 Swiss franc), subtraction in a problem solving context (picture of a toy with a price tag, picture with a bill: "You pay with the bill. How much change do you get?") and basic addition and subtraction (e.g.,  $4 + 4$ ;  $9 - 3$ ). WLE reliability for the whole test was .84 ( $N = 1,015$ ). WLE reliability for the seven addition and subtraction items was .73 ( $N = 1,015$ ).

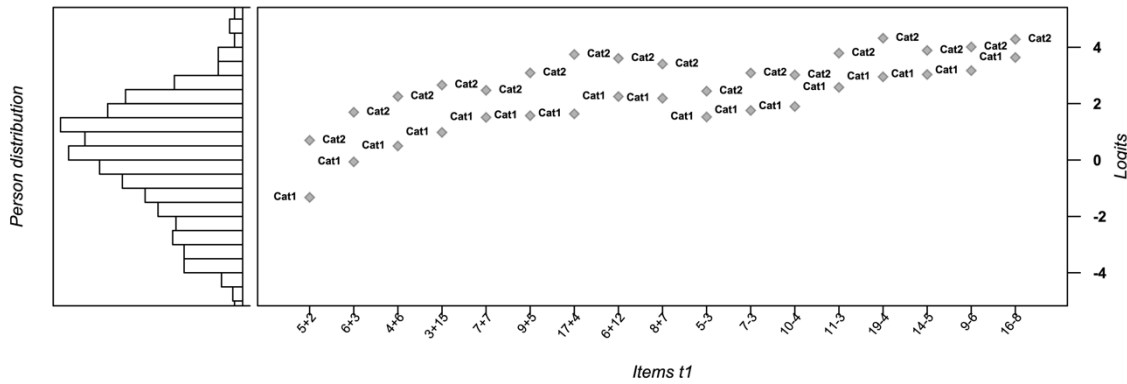
## Results

### Parameters of the CC Test (RQ 1)

All of the computation tasks were scored as partial credit items (1 point for a correct answer, cat1, an additional point for an NC strategy, cat2; see section [Development of the Instrument](#)). The easiest item was  $5 + 2$  (-0.31), the most difficult item was  $16 - 8$  (3.96; [Table 2](#)). [Figure 1](#) depicts the Wright Map for t1. Student scores were normally distributed. As expected, the test was difficult at the beginning of Grade 1, and only two of the children solved all of the problems. Most of the items had above an average difficulty of 0 ([Figure 1](#)). [Figure 1](#) shows that the cat1 score is always below the cat2 score. This means that the probability of a correct solution was always higher than the probability of using an NC strategy. Results show that the test was less difficult at t2 than at t1. At t2, most of the items were around or below the mean of 0 ([Table 2](#), [Figure 2](#)). As for t1, the t2 cat1 scores are always below the cat2 scores. Therefore, the probability of a correct solution was always higher than for using a NC strategy at t2. The easiest item was  $5 + 2$  (-2.86), the most difficult item was the new item  $20 - 13$ . The Wright Map ([Figure 2](#)) shows that student scores were normally distributed. Fifty-two children solved all of the problems at t2. The distribution of the items in the Wright Map also indicates that the person parameter fit to item difficulty is good – that all items could be solved by at least some of the children. As expected, this fit is worse at the beginning of the school year before the formal maths teaching.

**Figure 1**

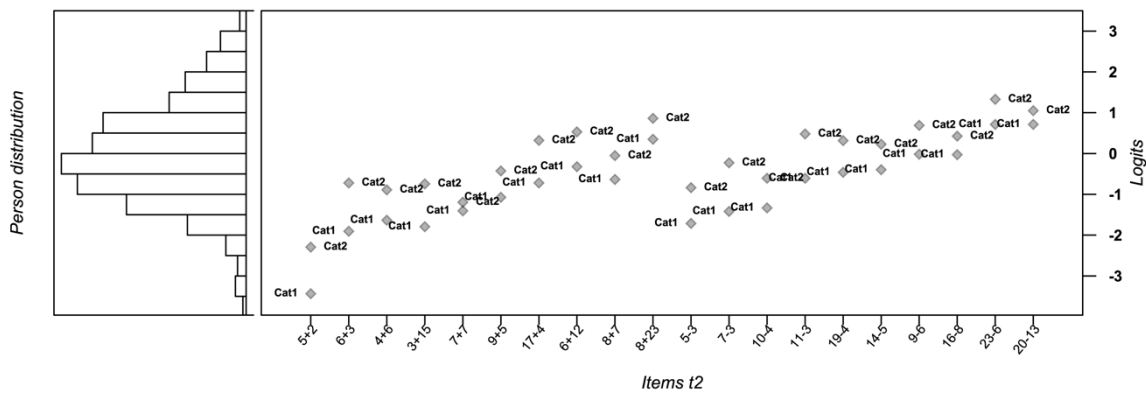
Wright Map of the CC Test at t1



Note. Cat1 = Category 1 (from score 0 to 1); Cat2 = Category 2 (from score 1 to 2).

**Figure 2**

Wright Map of the CC Test at t2



Note. Cat1 = Category 1 (from score 0 to 1); Cat2 = Category 2 (from score 1 to 2). The infit MNSQ for 20 items ranged from 0.88 to 1.16 (Table 2). Thus, all items had infit MNSQ values within the range of tolerance at both measurement points (Wilson, 2005).

To assess measurement invariance over time, DIF analyses of item difficulty were conducted (see section Analyses). Only one item (6 + 3) had a high negative DIF (logits -0.898 > 0.639) showing it to be easier over time relative to the rest of the items. Because we only had two single digit items with a result < 10, where the probability of using an NC strategy is high, the item was retained. To address the problem of measurement variance over time, the item with high negative DIF was treated as two separate items at t1 and t2 and the item parameter at t2 was estimated independently of that at t1. That is, the item was part of the scale within measurement points, but not anchored between measurement points. Therefore, the item difficulties of this item were not equalized. Sixteen items were used to anchor the two measurement points. WLE reliability was acceptable at t1 (.77) and good at t2 (.87). Person ability scores ranged from -3.81 to 4.50 ( $M_{\text{LOGIT}} = -1.35, SD = 1.72$ ) at t1 and from -3.70 to 4.61 ( $M_{\text{LOGIT}} = 1.39, SD = 1.39$ ) at t2.



**Table 2***Item Difficulty, MNSQ, and Threshold Parameters of CC at t1 and t2*

Item	Item difficulty <sup>a</sup>		MNSQ		Thresholds <sup>c</sup> t1		Thresholds <sup>c</sup> t2	
	t1	t2	t1	t2	Cat 1 <sup>b</sup>	Cat 2 <sup>b</sup>	Cat 1	Cat 2
5 + 2	-0.31	-2.86	1.11	1.08	-1.33	0.70	-3.44	-2.29
6 + 3	0.81	-1.31	1.12	1.05	-0.07	1.70	-1.91	-0.72
4 + 6	1.38	-1.26	1.19	1.01	0.49	2.26	-1.64	-0.88
3 + 15	1.82	-1.27	1.14	1.05	0.97	2.67	-1.79	-0.74
7 + 7	1.99	-1.30	1.12	1.16	1.50	2.48	-1.41	-1.19
9 + 5	2.33	-0.75	1.04	1.07	1.57	3.10	-1.07	-0.42
17 + 4	2.69	-0.20	0.94	1.05	1.63	3.75	-0.73	0.33
6 + 12	2.92	0.10	0.96	0.98	2.24	3.61	-0.33	0.53
8 + 7	2.79	-0.34	0.92	1.02	2.18	3.41	-0.64	-0.05
8 + 23		0.60		0.94			0.34	0.87
5 - 3	1.98	-1.27	1.05	1.07	1.52	2.45	-1.71	-0.83
7 - 3	2.42	-0.82	1.03	1.06	1.75	3.09	-1.43	-0.22
10 - 4	2.45	-0.97	0.89	0.95	1.90	3.02	-0.34	-0.61
11 - 3	3.18	-0.06	1.04	1.13	2.57	3.80	-0.61	0.48
19 - 4	3.64	-0.07	0.90	0.94	2.95	4.33	-0.46	0.32
14 - 5	3.46	-0.08	0.93	0.93	3.03	3.89	-0.40	0.24
9 - 6	3.59	0.34	0.94	0.99	3.17	4.01	-0.02	0.70
16 - 8	3.96	0.20	0.83	0.94	3.64	4.29	-0.03	0.43
23 - 6		1.02		0.98			0.71	1.34
20 - 13		0.88		0.88			0.72	1.05

<sup>a</sup>Item difficulty in the Partial Credit Model is displayed as logits (log odds unit). <sup>b</sup>Category 1 (from score 0 to 1) and Category 2 (from score 1 to 2).

<sup>c</sup>Thurstonian Thresholds.

### Criterion-Related Validity

The criterion-related validity analysis examined the degree to which measurements of CC and NC strategy use at the beginning of Grade 1 correlated with the numeracy test and the addition and subtraction part of this test. A correlation coefficient of  $r = .65$  ( $p < .01$ ,  $N = 1,015$ ) was found for the overall test. The correlation of the CC test with the addition and subtraction part (7 items) was  $.73$  ( $p < .01$ ,  $N = 1,015$ ).

### Relationship Between a Correct Result and NC Strategy Use (RQ 2)

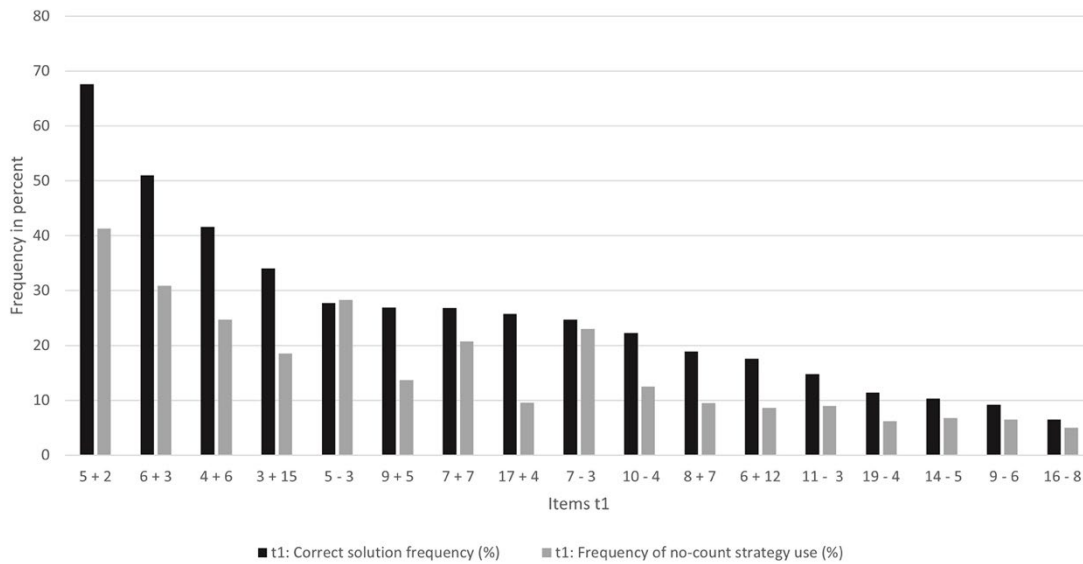
To determine whether a child's choice of strategy depends on the difficulty of the math problem or their familiarity with NC strategies learned during Grade 1, we performed an analysis to check if the combined assessment of correctness and NC strategy use in a three-level variable (0, 1, 2) is valid.

In a first step, the percentage of participants who gave correct answers for each problem and the percentage of participants who used an NC strategy to solve each problem were tabulated independently of each other at t1 (Figure 3) and at t2 (Figure 4). The problems were arranged in order of increasing difficulty, from left to right. Figure 3 shows that NC strategies were already being used to solve problems with varying levels of difficulty (minimum value 5%, maximum value 40%) at the beginning of Grade 1 (t1). The frequency of NC strategy use was highly correlated with the frequency of correct answers ( $r = .92$ ,  $p < .01$ ,  $n = 17$ ).

Figure 4 shows that at t2 (the end of Grade 1) one third of the children were solving more difficult problems, such as 20 - 13, without counting. At least 50% of the children used NC strategies to solve nine of the 20 problems. The correlation between the frequency of NC strategies and the frequency of correctness ( $r = .95$ ,  $p < .01$ ,  $n = 20$ ) was also high at t2.

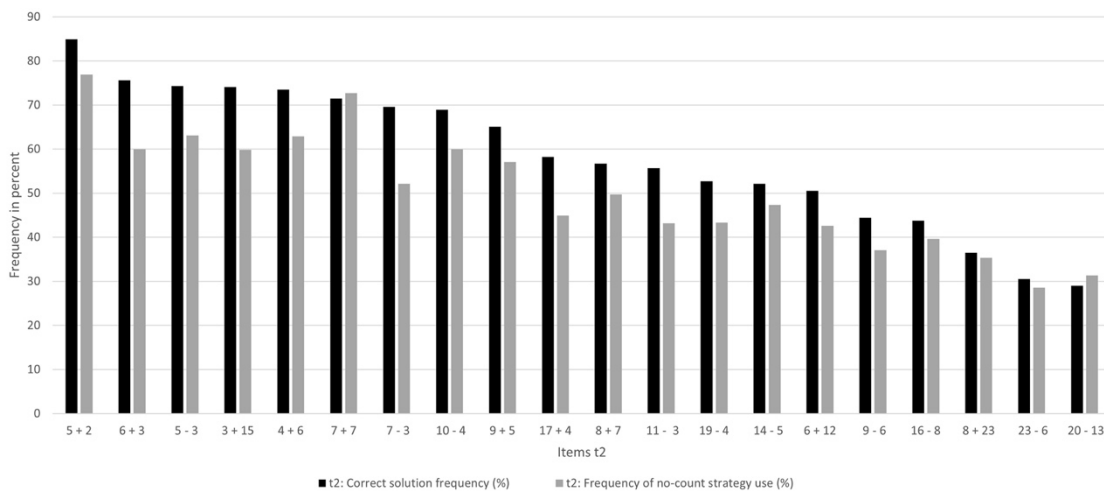
**Figure 3**

*Correct Solution Frequency (%) and NC Strategy Use per Math Problem (%) at t1*



**Figure 4**

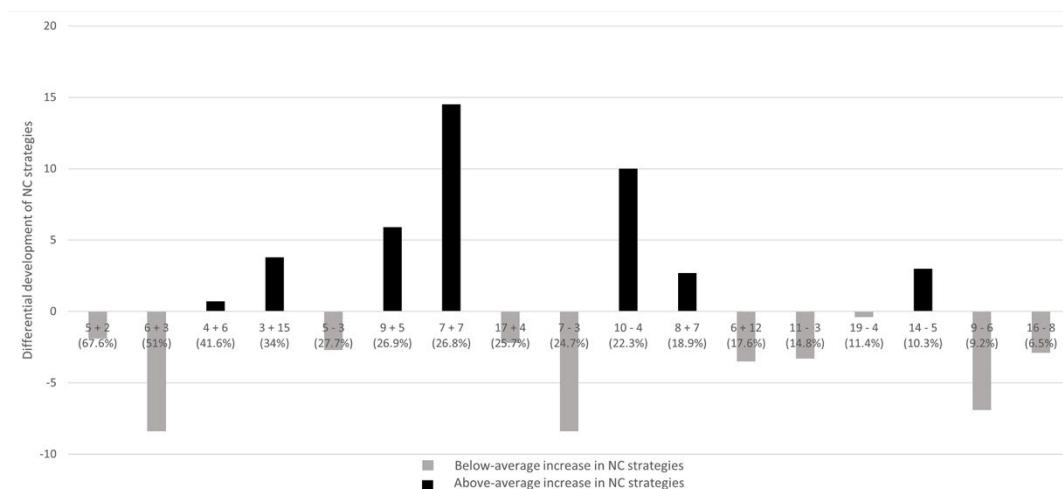
*Correct Solution Frequency (%) and NC Strategy Use per Math Problem (%) at t2*



In a second step, the development of NC strategy use from t1 to t2 was examined. In Figure 5 the difficulty of the problem and the increased use of NC strategies were visualized independently of each other. The item-specific differences in the increase in the use of NC strategies from t1 to t2 were calculated with a mean difference across all items. For each problem, the difference between the frequency of NC strategies at t1 and t2 was calculated. The mean value of all difference values was then subtracted from each single difference value. The resulting values are shown in Figure 5, ordered by difficulty at t1. It reveals whether NC strategies increased at an average (at the zero point), below- or above-average rate.

Figure 5

Differential Development of NC Strategies as Related to Problem Difficulty at t1



An above-average increase in NC strategies was found for problems associated with the *number 10* ( $4 + 6$ ,  $10 - 4$ ), *power of five* ( $3 + 15$ ,  $9 + 5$ ,  $14 - 5$ ), *doubling* ( $7 + 7$ ), and *doubling + 1* ( $8 + 7$ ) (highlighted in black). An above-average increase in NC strategies was even observed for difficult problems, such as  $8 + 7$  and  $14 - 5$ , which were correctly solved by less than 20% of the children at t1. A below-average increase was found for problems which were more frequently solved using an NC strategy at t1 ( $5 + 2$ ,  $6 + 3$ ,  $5 - 3$ ,  $7 - 3$ ). However, a below-average increase in NC strategies was particularly evident in subtraction problems without a specific structure (e.g.,  $11 - 3$ , no relation to the power of five).

Finally, the threshold parameters (Table 2) were analyzed to examine the difficulty of crossing the boundary between scoring 0 and scoring 1, and between scoring 1 and 2. At t1, all thresholds from 0-1 (from false to correct with a counting strategy) were below the corresponding 1-2 thresholds (from correct with a counting strategy to correct with an NC strategy). At t2 the 0-1 thresholds were all also below the 1-2 thresholds. This shows that it was also more difficult for the children to solve a task correctly at t2 using an NC strategy rather than a counting strategy. The 1-2 thresholds were lower at the end of Grade 1 than at the beginning. This is evidence that at t2 many of the correctly solved tasks were solved using an NC strategy. The step from a correct solution with a counting strategy to correct solution with an NC strategy seems to have been smaller at t2.

The ranges of the 0-1 and the 1-2 thresholds overlapped at both measurement points. This is evidence that the ability to correctly solve computation problems and to do so without counting seem to coexist in first graders.

## Discussion

In this paper we presented a new instrument for assessing any change/difference in CC and NC strategy use by students between the beginning and end of Grade 1. The instrument was designed to ensure measurement invariance over the course of Grade 1. The relationship between item difficulty and strategy use was analyzed using the partial credit model (Wright & Masters, 1982).

### Assessing Computation Competence and NC Strategy Use (RQ 1)

The instrument was used to separately assess whether an answer was correct and whether a counting strategy had been used (counting strategy vs. decomposition or retrieval strategy). A combination of observation, reaction time, verbal statement, and a dual task method with tapping was used to discriminate counting strategies from NC ones. All items had MNSQ values in the tolerated range, which means that item homogeneity can be assumed. WLE reliability was

acceptable for t1 and good for t2. This suggests that the combination of different approaches, including the tapping to detect strategies, is a reliable assessment method for Grade 1.

It may be that the tapping triggers the use of a counting strategy. However, though the use of counting strategies was high for some items, the percentage of NC strategies recorded at t1 was higher than that in the study by Gaidoschik (2012). This means that children who were able to find a result by decomposition or retrieval used this strategy and were not prompted to count by the tapping. The tapping does appear to enable the observation of NC strategies more precisely than if there were no second task. Gaidoschik (2012), who conducted interviews, categorized 10.1 to 26% of strategies per item as “I do not know” at the beginning of Grade 1 (except for very easy items such as doubling 2). In this study, fewer (3 – 15%) of the strategies could not be assessed at t1.

Because children make large achievement gains in their first year of school, developing an assessment instrument that can be used at both the beginning and end of the year poses a challenge. This is also evident from our data. At the beginning of Grade 1, the test was difficult. At t2, the test was too easy for 52 students, who solved all items. Because the focus of this study is on low achievers, this is not an issue here. In addition, for measurement invariance over time, only one item had longitudinal DIF and failed to measure the same construct at t1 and t2. Therefore, this item was used to estimate child ability separately across the two measurement points without equating item difficulty. Given the children’s large learning gains during first grade, the results of the DIF analyses are very satisfactory. Finally, the probability of solving a problem correctly was always higher than the probability of using a NC strategy.

Criterion validity was evaluated using a group test that assessed numeracy and addition and subtraction competence (without considering strategy use). The correlation of overall numeracy test with the CC test (.65) was lower than the correlation of the addition and subtraction part of this test with the CC test (.73). We explain this difference by noting that the numeracy test assessed a broad range of informal mathematical competences (e.g., counting) and included only few addition and subtraction items, while the CC test only examined addition and subtraction. Other studies comparing the results of different mathematical tests at the beginning of Grade 1 also report correlations in the range of .70 (Ennemoser et al., 2017). Criterion validity is therefore satisfactory.

## Relationship Between a Correct Result and NC Strategy Use (RQ 2)

When testing CC and NC strategy use one needs to know if the child’s choice of strategy, counting vs. non-counting, is dependent on item difficulty (Reed et al., 2015; Siegler, 1987). In other words, we need to know if, at t1 the children had the ability to solve problems using an NC strategy. The descriptive analyses and the threshold analyses revealed that NC strategies were already being used regularly at t1. Moreover, the analyses confirm that the two aspects (correct result, NC strategy) can be classified on the same scale (substantial overlap of the range of the 0-1 thresholds with the 1-2 thresholds). This is also supported by the high correlation between the two manifest aspects of solution frequency and frequency of NC strategy use at t1 and t2.

The descriptive analyses show that the increase in NC strategies from t1 to t2 does not depend on the difficulty of the problem alone. Problems with a medium difficulty show the largest increase in NC strategies. The differential development in problems with medium difficulty can be explained by the characteristics of these problems. Due to their structure (power of five, decomposition of tens, doubling), they are particularly suited to the use of NC strategies. These kinds of problems are also often found in the textbooks used by the participating classes. More difficult problems tended to have a smaller, but still substantial, increase in NC strategies. In contrast, the easiest problems did not show the greatest increase. This can be explained by the fact that the proportion of NC was already high at the beginning of Grade 1.

The differential development of NC strategies as a function of problem difficulty at the beginning of Grade 1 thus highlights the complex interaction of problem characteristics and person characteristics (or NC strategy use).

## Limitations

The study has limitations. Some children were unable to solve computation problems at the beginning of Grade 1. To avoid frustration, there was a pre-test for CC and NC strategy use that asked the children to evaluate the correctness of addition problems. This differed from the format of the actual test where children had to derive and state the result on

their own and therefore may have affected the results. It was also assumed that a child who was unable to solve three consecutive problems would be unable to solve the rest, which were scored 0. According to probabilistic testing theory, children who cannot solve a problem may also randomly give a correct answer. Our procedure created extreme cases where some students scored 0, which affected the assumption of normality. This limitation had to be accepted for the following reasons: First, it was important to avoid frustrating children who were unable to perform CC. Second, shorter tests are more economical (time, financial resources, Pospeschill, 2010). Third, participants with missing scores could not be used in subsequent longitudinal analyses.

First graders' language skills can also pose problems during assessment because younger children have more limited vocabularies. When asked "How did you figure it out?" some children responded, "I don't know." It was not always clear if they were unable to calculate the problem in their head or if they were simply unable to explain their calculation strategy. This could be a common problem when assessing strategy use.

Only one person scored strategy use for each child. Calculating interrater reliability would therefore have required the analysis of video recordings of the assessments. Video analyses of over 1.000 students would have been extremely labor intensive and requesting permission to record the students in the in-depth examination may have reduced our recruitment rate. However, the test administrators were given rigorous training and if strategy use was unclear, it was assessed as missing. A further limitation is that we did not analyze the NC strategy use of incorrectly solved problems because the aim of the study was to develop an instrument to assess counting and NC strategies. Such analyses would give important insight into the development of CC and NC strategy use in young children.

## Conclusions

The results of this study are promising. The tool can be used in its present form to assess CC and NC strategy use at the beginning and end of Grade 1. Nevertheless, the question of how to capture the use of NC strategies by first graders accurately and efficiently requires more research. Future studies should pay particular attention to NC strategies used in incorrectly solved problems. Since the verbalization of computation strategies is a challenge for younger children, follow-up studies could investigate whether eye tracking might be a suitable method for assessing strategy use in this age group (Schindler & Lilienthal, 2018).

---

**Funding:** This study was funded by swissuniversities (P-9/TP2)

---

**Acknowledgments:** The authors have no additional (i.e., non-financial) support to report.

---

**Competing Interests:** The authors also confirm that there are no known conflicts of interest associated with this publication. There was no significant financial support for this work that could have influenced its outcome. All authors read and approved the final manuscript.

---

**Ethics Statement:** The study has been carried out in accordance with relevant ethical principles and standards. The project was approved by the ethical committee of the Faculty of Arts and Social Sciences of the University of Zurich. Teachers voluntarily chose to participate in the study and children had written parental consent.

---

**Related Versions:** The study is part of a larger project. The article reports selected results from a dissertation that is published on an internal repository (Leuenberger, 2021).

---

**Data Availability:** Data are available upon request to the corresponding author.

---

## References

- Aunio, P., Heiskari, P., van Luit, J. E., & Vuorio, J.-M. (2015). The development of early numeracy skills in kindergarten in low-, average- and high-performance groups. *Journal of Early Childhood Research, 13*(1), 3–16.  
<https://doi.org/10.1177/1476718X14538722>

- Baroody, A. J., Purpura, D. J., Eiland, M. D., & Reid, E. E. (2014). Fostering first graders' fluency with basic subtraction and larger addition combinations via computer-assisted instruction. *Cognition and Instruction, 32*(2), 159–197. <https://doi.org/10.1080/07370008.2014.887084>
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates.
- 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>
- Clements, D. H., & Sarama, J. (2007). Early childhood mathematics learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 461–555). Information Age.
- Cowan, R. (2003). Does it all add up? Changes in children's knowledge of addition combinations, strategies, and principles. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructive adaptive expertise* (pp. 35–74). Lawrence Erlbaum Associates.
- De Smedt, B., Torbeyns, J., Stassens, N., Ghesquière, P., & Verschaffel, L. (2010). Frequency, efficiency and flexibility of indirect addition in two learning environments. *Learning and Instruction, 20*(3), 205–215. <https://doi.org/10.1016/j.learninstruc.2009.02.020>
- Deutscher, T. (2012). *Arithmetische und geometrische Fähigkeiten von Schulanfängern* [Arithmetical and geometric competences of children at the beginning of Grade 1]. Vieweg und Teubner Verlag. <https://doi.org/10.1007/978-3-8348-8334-6>
- Ennemoser, M., Krajewski, K., & Sinner, D. (2017). *MBK 1+. Test mathematischer Basiskompetenzen ab Schuleintritt* [Assessing basic mathematical competence at the beginning of Grade 1]. Hogrefe.
- Gaidoschik, M. (2012). First-graders' development of calculation strategies: How deriving facts helps automatize facts. *Journal für Mathematik-Didaktik, 33*(2), 287–315. <https://doi.org/10.1007/s13138-012-0038-6>
- Grube, D., & Seitz-Stein, K. (2012). Arbeitsgedächtnis und Rechnen [Working memory and computation]. In M. Hasselhorn & C. Zoelch (Eds.), *Funktionsdiagnostik des Arbeitsgedächtnisses* (pp. 145–157). Hogrefe.
- Hartig, J., & Kühnbach, O. (2006). Schätzung von Veränderung mit „plausible values“ in mehrdimensionalen Rasch-Modellen [Estimation of change with “plausible values” in multidimensional Rasch models]. In A. Ittel & H. Merckens (Eds.), *Veränderungsmessung und Längsschnittstudien in der empirischen Erziehungswissenschaft* (pp. 27–44). VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-90502-0\\_3](https://doi.org/10.1007/978-3-531-90502-0_3)
- Häsel-Weide, U. (2016). *Vom Zählen zum Rechnen. Struktur-fokussierende Deutungen in kooperativen Lernumgebungen* [From counting to arithmetic: Structure-focused interpretations in cooperative learning environments]. Springer Spektrum.
- Hopkins, S., & Bayliss, D. (2017). The prevalence and disadvantage of min-counting in seventh grade: Problems with confidence as well as accuracy? *Mathematical Thinking and Learning, 19*(1), 19–32. <https://doi.org/10.1080/10986065.2017.1258613>
- Hopkins, S., Russo, J., & Siegler, R. (2022). Is counting hindering learning? An investigation into children's proficiency with simple addition and their flexibility with mental computation strategies. *Mathematical Thinking and Learning, 24*(1), 52–69. <https://doi.org/10.1080/10986065.2020.1842968>
- Jordan, N. C., Hanich, L. B., & Uberti, H. Z. (2003). Mathematical thinking and learning difficulties. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills* (pp. 359–383). Lawrence Erlbaum Associates.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up*. National Academies Press. <https://doi.org/10.17226/9822>
- Leuenberger, D. (2021). *Erfassung und Lernzuwachs von nicht-zählenden Kopfrechenkompetenzen im ersten Schuljahr: Eine Interventionsstudie zur Ablösung vom zählenden Rechnen im inklusiven Unterricht* [Assessing learning gains of non-count mental arithmetic skills in the first year of school: An intervention study on disengagement from counting strategies in inclusive education] [Doctoral dissertation, University of Zurich]. ZORA (Zurich Open Repository and Archive). <https://doi.org/10.5167/uzh-208222>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. <https://doi.org/10.1007/BF02296272>
- Morano, S., Randolph, K., Markelz, A. M., & Church, N. (2020). Combining explicit strategy instruction and mastery practice to build arithmetic fact fluency. *Teaching Exceptional Children, 53*(1), 60–69. <https://doi.org/10.1177/0040059920906455>
- Moser Opitz, E. (2001). Mathematical knowledge and progress in the mathematical learning of children with special needs in their first year of school. In M. Van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 207–210). PME.
- Moser Opitz, E. (2013). *Rechenschwäche – Dyskalkulie. Theoretische Klärungen und empirische Studien an betroffenen Schülerinnen und Schülern* [Dyscalculia: Theoretical clarifications and empirical studies on affected pupils]. Haupt.

- Paek, I., & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel–Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement, 71*(6), 1023–1046. <https://doi.org/10.1177/0013164411400734>
- Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation: Mit 77 Fragen zur Wiederholung* [Theory, construction, and evaluation of tests: 77 practice questions]. Ernst Reinhardt Verlag.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research.
- Reed, H. C., Stevenson, C., Broens-Paffen, M., Kirschner, P. A., & Jolles, J. (2015). Third graders' verbal reports of multiplication strategy use: How valid are they? *Learning and Individual Differences, 37*, 107–117. <https://doi.org/10.1016/j.lindif.2014.11.010>
- Robitzsch, A., Kiefer, T., & Wu, M. (2020). *Package TAM. Test Analysis Modules* (R package version 3. 5–19) [Computer software]. <https://cran.r-project.org/package=TAM>
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2nd ed.) [Course book theory and construction of tests]. Huber.
- Schindler, M., & Lilienthal, A. J. (2018). Eye-tracking for studying mathematical difficulties—Also in inclusive settings. In: *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 115–122). PME.
- Schwab, S., & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen [Testing for measurement invariance in students with and without special educational needs – A case example using the short form of the Illinois Loneliness and Social Satisfaction Scale]. *Empirische Sonderpädagogik, 7*(3), 175–193. <https://doi.org/10.25656/01:11380>
- Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405–410. <https://doi.org/10.1111/1467-9280.00076>
- Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*(3), 250–264. <https://doi.org/10.1037/0096-3445.116.3.250>
- Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Lawrence Erlbaum Associates.
- Sievert, H., van den Ham, A.-K., & Heinze, A. (2021). Are first graders' arithmetic skills related to the quality of mathematics textbooks? A study on students' use of arithmetic principles. *Learning and Instruction, 71*, Article 101401. <https://doi.org/10.1016/j.learninstruc.2020.101401>
- Vanbinst, K., Ghesquière, P., & De Smedt, B. (2014). Arithmetic strategy development and its domain-specific and domain-general cognitive correlates: A longitudinal study in children with persistent mathematical learning difficulties. *Research in Developmental Disabilities, 35*(11), 3001–3013. <https://doi.org/10.1016/j.ridd.2014.06.023>
- Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology, 6*, Article 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- Vasilyeva, M., Laski, E. V., & Shen, C. (2015). Computational fluency and strategy choice predict individual and cross-national differences in complex arithmetic. *Developmental Psychology, 51*(10), 1489–1500. <https://doi.org/10.1037/dev0000045>
- Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Information Age.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. <https://doi.org/10.1007/BF02294627>
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.
- Wittich, C. (2017). *Mathematische Förderung durch kooperativ-strukturiertes Lernen: Eine Interventionsstudie zur Ablösung vom zählenden Rechnen an Grund- und Förderschulen* [Mathematical support through cooperative-structured learning: An intervention study on non-counting strategies at primary and special schools]. Springer.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Mesa Press.



*Journal of Numerical Cognition* (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.