

Number Line Tasks and Their Relation to Arithmetics in Second to Fourth Graders

Carrie Georges^a, Christine Schiltz^a

[a] *Institute of Cognitive Science and Assessment (COSA), Department of Behavioural and Cognitive Sciences (DBCS), Faculty of Humanities, Education and Social Sciences (FHSE), University of Luxembourg, Esch-Belval, Luxembourg.*

Journal of Numerical Cognition, 2021, Vol. 7(1), 20–41, <https://doi.org/10.5964/jnc.6067>

Received: 2020-07-24 • Accepted: 2021-01-04 • Published (VoR): 2021-03-31

Corresponding Author: Carrie Georges, Institute of Cognitive Science and Assessment (COSA), Department of Behavioural and Cognitive Sciences (DBCS), Faculty of Humanities, Education and Social Sciences (FHSE), University of Luxembourg, Maison des Sciences Humaines, 11, Porte des Sciences, L-4366 Esch-Belval, Luxembourg. E-mail: carrie.georges@uni.lu

Abstract

Considering the importance of mathematical knowledge for STEM careers, we aimed to better understand the cognitive mechanisms underlying the commonly observed relation between number line estimations (NLEs) and arithmetics. We used a within-subject design to model NLEs in an unbounded and bounded task and to assess their relations to arithmetics in second to fourth grades. Our results mostly agree with previous findings, indicating that unbounded and bounded NLEs likely index different cognitive constructs at this age. Bounded NLEs were best described by cyclic power models including the subtraction bias model, likely indicating proportional reasoning. Conversely, mixed log-linear and single scalloped power models provided better fits for unbounded NLEs, suggesting direct estimation. Moreover, only bounded but not unbounded NLEs related to addition and subtraction skills. This thus suggests that proportional reasoning probably accounts for the relation between NLEs and arithmetics, at least in second to fourth graders. This was further confirmed by moderation analysis, showing that relations between bounded NLEs and subtraction skills were only observed in children whose estimates were best described by the cyclic power models. Depending on the aim of future studies, our results suggest measuring estimations on unbounded number lines if one is interested in directly assessing numerical magnitude representations. Conversely, if one aims to predict arithmetic skills, one should assess bounded NLEs, probably indexing proportional reasoning, at least in second to fourth graders. The present outcomes also further highlight the potential usefulness of training the positioning of target numbers on bounded number lines for arithmetic development.

Keywords

number line estimation, arithmetics, mathematical development, numerical cognition, elementary school children, psychophysical function

The Importance of NLE for Arithmetics

To prepare students for success in STEM careers, it is important to better understand the cognitive processes underlying mathematical development. Several basic numerical tasks, such as magnitude comparison (e.g., [Schneider et al., 2017](#)) or number line estimation (e.g., [Schneider et al., 2018](#)) have been particularly successful to predict children's math skills and thus improve our understanding of mathematical learning processes.

Studies consistently report strong relations between children's number line estimations (NLEs) and arithmetics. Better estimations in terms of average error have been related to higher achievements on various more complex and advanced mathematical skill measures in elementary school children (e.g., [Booth & Siegler, 2006](#); [Fazio, Bailey, Thompson, & Siegler, 2014](#); [Geary, 2011](#); [Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013](#); [Siegler & Booth, 2004](#); see [Siegler, 2016](#), for a review). Importantly, relations to arithmetics remained significant even after controlling for



potentially confounding variables, such as socio-economic status, working memory, intelligence, reading achievement or non-symbolic numerical knowledge (Geary, 2011; Hansen et al., 2015; Hornung, Schiltz, Brunner, & Martin, 2014; Vukovic et al., 2014).

In NLE tasks, participants need to place a target number on a visually presented number line (Siegler & Opfer, 2003). In the classical “bounded” version of the task, number lines are labelled with two endpoints (Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014). More recently, an “unbounded” task has been introduced, where only the origin of the number line and a scaling unit (e.g., 0-1) are given (Cohen & Blanc-Goldhammer, 2011).

Since NLE tasks were typically considered to allow for a direct assessment of the underlying mental number line (MNL) representation (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Booth & Siegler, 2006, 2008; Siegler & Opfer, 2003), associations between NLEs and arithmetics have been widely regarded as evidence for a critical role of numerical magnitude representations in arithmetic development (Schneider, Grabner, & Paetsch, 2009; Schneider et al., 2008; Siegler & Opfer, 2003). Accordingly, the ability to represent and process numerical magnitudes likely supports the acquisition of more advanced mathematical skills. The beneficial role of the MNL is not only empirically supported (e.g., Georges, Hoffmann, & Schiltz, 2017; Hoffmann, Hornung, Martin, & Schiltz, 2013), but also a key assumption in several dominant theories of mathematical learning and development, including the integrated theory of numerical development (Siegler, 2016; Siegler & Lortie-Forgues, 2014; Siegler, Thompson, & Schneider, 2011), Dehaene’s (1997) account of number sense, and the theory of mathematical core knowledge (e.g., Feigenson, Dehaene, & Spelke, 2004).

However, recent studies assessing both the classical bounded NLE task and the relatively new unbounded version found that associations between NLEs and arithmetics were only observed for the former but not the latter task (Jung et al., 2020; Link, Nuerk, & Moeller, 2014; but see Kim & Opfer, 2017). This finding also agrees with the meta-analysis of Schneider and colleagues (2018), where correlations with mathematical skills were moderated by the variant of the NLE task. More concretely, relations with mathematical competence were significantly positive for the standard, bounded version of the NLE task but not significantly different from zero for unbounded number lines. This not only suggests that both tasks likely reflect different cognitive constructs, but also questions whether numerical magnitude representations can truly account for the relation between bounded NLEs and arithmetics.

The Determinants of NLE

Interestingly, bounded NLEs seem to be best described by a variety of cyclic power models (Hollands & Dyre, 2000), supposedly capturing calculation strategies, such as proportional judgement, rather than direct estimation (Barth & Paladino, 2011; Barth et al., 2016; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder & Geary, 2014; Slusser, Santiago, & Barth, 2013; but see Opfer, Thompson, & Kim, 2016; Kim & Opfer, 2017), thus suggesting that calculation strategies as opposed to numerical magnitude representations probably underlie relations to arithmetics.

More concretely, since bounded number lines have a clearly defined origin and endpoint, individuals can scale the line length to a proportion by estimating the distances of the target number from the left and right boundaries of the line. For example, if individuals need to place the number 60 on a line from 0 to 100, they pick a position on the line and then look back and forth between that position and the boundaries of the line, adjusting the position until its distances to the lower and upper boundaries seem to be 60 and 40 respectively. When participants use this “proportional judgement” strategy to estimate the position of numbers on bounded lines, their estimates form an S-shaped ogival curve around the accuracy line, which can be described by Spence’s power model (Spence, 1990; Spence & Krizel, 1994). Hollands and Dyre (2000) adapted the power model to include the possibility of using multiple reference points. This so called “cyclic power model (CPM)” predicts multiple cycles of the S-shaped ogival pattern, depending on the number of reference points. Some individuals might only rely on the origin and endpoint, reflected by a single cycle (i.e., one-cycle power model, 1CPM), while others might use the middle of the line as an additional reference point, yielding two cycles of the S-shaped ogival pattern (i.e., two-cycle power model, 2CPM). Cohen and Sarnecka (2014) later introduced a variation to the CPM, the subtraction bias cyclic model (SBCM). This model is identical to the 1CPM but includes a subtraction bias, which indexes difficulties in the “subtract-and-compare” process. This model was shown to describe the estimates of younger children, suggesting that poor mensuration skills underlie the negatively accelerating (logarithmic-like) pattern of data usually produced by younger children on the bounded task (Cohen & Sarnecka,

2014). However, it should be noted that although CPMs can be indicative of the application of proportional reasoning, the latter calculation strategy is not necessarily exclusively captured by the superior fits of those power functions. Namely, Yuan, Prather, Mix, and Smith (2020) have recently shown that power models provided relatively poor fits to performances on a clustered dot array task that is readily solvable by basic proportional reasoning skills. Thus, although power models likely index the reliance on proportional judgement, poor fits do not necessarily suggest the absence of such a calculation strategy.

Even though the question of whether proportional reasoning is adequately modelled by CPMs remains a matter of debate, the application of such a calculation strategy on bounded tasks can be further supported by studies focussing on participants' error rates (e.g., Ashcraft & Moore, 2012). Namely, smaller error variability is typically observed close to the mid- and endpoints of the number line, resulting in a characteristic M-shaped distribution, thereby suggesting the reliance on reference points as a guidance for placing estimates on bounded lines (Cohen & Blanc-Goldhammer, 2011; Link, Huber, Nuerk, & Moeller, 2014). The use of proportional judgement is further corroborated by participants' verbal reports and observations of their solution behavior (Peeters, Degrande, Ebersbach, Verschaffel, & Luwel, 2016; Peeters, Verschaffel, & Luwel, 2017; Petitto, 1990), as well as eye-movement data (Schneider et al., 2008; Sullivan, Juhasz, Slattery, & Barth, 2011).

As opposed to the bounded task, strategies on unbounded number lines were shown to consist of direct estimation and dead-reckoning (Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Jung et al., 2020; Link, Huber, et al., 2014; Reinert, Hartmann, Huber, & Moeller, 2019; Reinert, Huber, Nuerk, & Moeller, 2017). Namely, participants use the given unit to either directly estimate target numbers by moving straight to their position on the number line or apply dead-reckoning by firstly estimating a working window of a specific size [e.g., 5 units] and by then using multiples of this working window to estimate the location of larger numbers. Estimating the position of numbers by repeatedly counting to smaller numbers (i.e., estimating the line length of the working window) creates a scalloped pattern in participants' estimates. This pattern of estimation can be described by a simple variant of Stevens's power law, termed the scalloped power model (SPM; Cohen & Blanc-Goldhammer, 2011). When participants estimate numbers directly, their estimates are characterized by a single scallop (i.e., single-scalloped power model, 1SPM). Conversely, dual- (2SPM) and multi-scalloped power models (multi-SPM) reflect the estimates of individuals applying the working window once or multiple times respectively before counting towards the position of the estimate. Considering that the SPMs indicate the reliance on basic counting skills, such as repeated addition, rather than any advanced calculation strategy, NLEs best described by these functions are considered to provide a more accurate picture of numerical magnitude representations (see Cohen & Sarnecka, 2014).

Direct estimation on unbounded number lines is also further confirmed by participants' error data. Namely, error variability in the unbounded task linearly increases with target number and is not characterized by the M-shaped pattern reflecting proportional judgement on bounded number lines (Link, Huber, et al., 2014).

Considering that 1) mainly bounded but not unbounded estimates correlate with arithmetics and that 2) direct estimation is likely only indexed by unbounded NLEs, mechanisms other than numerical magnitude representations such as proportional judgement probably underlie the relation between NLEs and arithmetics. In other terms, NLEs correlate with mathematical skills, because proportional reasoning is a key proficiency component in mathematics (cf., Boyer, Levine, & Huttenlocher, 2008), but not because numerical magnitude representations are an important determinant of mathematical development. The unbounded NLE task is thought to provide a better measure of the MNL (Cohen & Sarnecka, 2014). However, the absence of a relation between arithmetics and unbounded NLEs (Link, Nuerk, et al., 2014) questions the importance of numerical magnitude representations for mathematical learning.

An Ongoing Debate

Whether or not the MNL can account for the commonly observed relation between NLEs and arithmetics is a matter of ongoing debate. Although abundant evidence suggests that bounded estimations rather reflect calculation strategies (Barth & Paladino, 2011; Barth et al., 2016; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder & Geary, 2014; Slusser et al., 2013), which then account for relations to arithmetics, these observations conflict with the recent findings of Kim and Opfer (2017; see also Opfer et al., 2016). Namely, they showed that a mixed log-linear

model (MLLM) produced superior fits to bounded estimates than a model integrating all variants of the CPMs, termed mixed cyclic power model (MCPM). Since the MLLM predicts estimates as a weighted sum of logarithmic and linear transforms of the target number, it reflects the idea of age-related log-to-linear shifts in numerical representations (Anobile, Cicchini, & Burr, 2012; Cicchini, Anobile, & Burr, 2014; Dehaene, Izard, Spelke, & Pica, 2008; Kim & Opfer, 2017; Opfer et al., 2016). Their findings therefore suggest that not only unbounded but also bounded estimates properly reflect numerical magnitude representations. Interestingly, what is more, the logarithmicity component of the MLLM was found to significantly relate to addition and subtraction skills. This indicates that the degree of logarithmic compression in numerical magnitude representations probably accounts for the relation between bounded NLEs and arithmetics. The importance of the MNL in explaining associations between NLEs and mathematical skills is further corroborated by their observation that even unbounded estimates (assumed to index numerical representations more directly) significantly correlated with arithmetics (Kim & Opfer, 2017). In sum, Kim and Opfer (2017) thus provide evidence in favour of the initial idea that the MNL underlies the commonly observed relation between NLEs and arithmetics.

How can the findings of Kim and Opfer (2017) be reconciled with evidence against the importance of the MNL in the relation between NLEs and arithmetics? One possible explanation for this discrepancy is that studies not finding a relation between unbounded NLEs and arithmetics were mainly conducted in relatively older elementary school children (Link, Nuerk, et al., 2014) or adolescents (Jung et al., 2020), while Kim and Opfer (2017) tested younger children, attending preschool, first or second grade. Notably, the importance of the MNL for unbounded NLEs has been confirmed in older children (Link, Huber, et al., 2014) and adolescents (Jung et al., 2020), suggesting that the reliance on numerical magnitude representations might well account for the lack of relation to arithmetics at these later developmental stages. This assumption has, however, not yet been confirmed with intra-individual studies in older elementary school children. More concretely, only studies focussing on either younger children (Kim & Opfer, 2017) or adolescents (Jung et al., 2020) have assessed the constructs underlying NLEs as well as their relations to arithmetics in the same individuals. It is thus advisable to employ such a within-subject approach also in slightly older elementary school children to determine whether relations between arithmetics and bounded but not unbounded NLEs previously observed in fourth graders (Link, Nuerk, et al., 2014) can actually be explained by the reliance on proportional judgement on bounded as opposed to the MNL on unbounded number lines in such older elementary school children (Link, Huber, et al., 2014).

Aim

In this paper, we generally aimed to further unravel the cognitive mechanisms underlying the commonly observed relation between NLEs and arithmetics. To shed further light on the aforementioned inconsistencies and to account for the potential shortcomings of previous designs, the present study used a *within-subject approach* to assess 1) model fits for unbounded and bounded NLEs (and as such their underlying cognitive determinants) as well as 2) their relations to arithmetics within the same individuals. We focussed on slightly *older elementary school children* attending second to fourth grade to complement previous studies using a similar within-subject approach in younger children (Kim & Opfer, 2017) or adolescents (Jung et al., 2020) and to get a better understanding of the mechanisms actually explaining the relation between bounded but not unbounded NLEs and arithmetics previously reported in fourth graders (Link, Nuerk, et al., 2014).

The current *hybrid design* (i.e., the use of a modelling procedure to better understand the constructs underlying bounded and unbounded NLEs and consequently individual differences in addition and subtraction skills, see e.g., Kim & Opfer, 2017, or Jung et al., 2020, for a similar design) enabled us to further assess the mechanisms explaining relations between NLEs and arithmetics without the hurdle of incorporating additional tests to determine the influence of potentially confounding domain-specific and domain-general variables previously shown to relate to NLEs and/or arithmetics (e.g., visuospatial abilities, see Sella, Sader, Lollot, & Cohen Kadosh, 2016; Simms, Clayton, Cragg, Gilmore, & Johnson, 2016).

We also performed *moderation analysis* to determine whether potential associations between NLEs and arithmetics depend on the best-fit model, thereby further unravelling the cognitive construct actually accounting for their relation. More concretely, we hypothesized that if bounded NLEs reflect proportional judgement, which in turn explains their

relation to mathematics, associations might be strongest in those children whose estimations are best described by functions capturing this kind of calculation strategy.

We intended to contrast the MNL with proportional reasoning as possible constructs underlying unbounded and bounded NLEs and their relations to arithmetics. Since the likelihood of strategy application and the reliable use of reference points was shown to be increased by employing a *familiar number range* (e.g., Slusser et al., 2013; White & Szűcs, 2012), we focussed on numbers in the range of 0-20 and 0-100 (depending on the task variant, see Link, Huber, et al., 2014, and Link, Nuerk, et al., 2014, for comparable differences in number ranges between bounded and unbounded tasks). This covers a familiar interval also in the youngest children of the present sample (e.g., Dackermann et al., 2018). The linear-like estimation patterns, usually produced by second to fourth graders on such familiar number lines (Booth & Siegler, 2006; Sella et al., 2015; Siegler & Booth, 2004), were then modelled using functions capturing either the MNL (i.e., the MLLM and SPMs) or calculation strategies, such as proportional reasoning (i.e., CPMs).

Considering the logarithmic-like (as opposed to linear-like) estimation patterns usually observed in younger children or with less familiar number ranges, we would like to point out that different conclusions regarding the constructs underlying (bounded and unbounded) NLEs and their relations to arithmetics might be drawn in those cases. It was, however, beyond the scope of the present study to examine whether the shift from seemingly logarithmic to more linear NLEs with age and experience reflects 1) changes in the disposition of numerical magnitude representations (e.g., Siegler & Opfer, 2003; Siegler & Booth, 2004), 2) the maturation of calculation strategies (e.g., the successful use of reference points, see Slusser et al., 2013), 3) the development of better mensuration skills (e.g., Cohen & Sarnecka, 2014) or 4) the reliance on different cognitive constructs (e.g., MNL and calculation strategies in younger and older individuals respectively, Dackermann et al., 2015) and how this shift might affect the relations to arithmetics across development. Given the linear-like NLEs of older children on familiar number lines, we did also not focus on any additional models, such as the bilinear account, suggested to provide good fits for logarithmic-like NLEs and thereby yet another alternative explanation for the logarithmic-to-linear shift hypothesis (see Ebersbach et al., 2008, or Moeller et al., 2009).

Method

Participants

A cross-sectional sample of 69 elementary school children (20 second graders [11 boys, mean age = 8.17 years, $SD = 0.37$], 18 third graders [9 boys, mean age = 9.32 years, $SD = 0.55$], 31 fourth graders [14 boys, mean age = 10.40 years, $SD = 0.46$]) was assessed on arithmetic and NLE tasks. All children participated voluntarily and were included in the sample only after their parents provided a signed informed consent form. The study was approved by the local Ethics Review Panel (ERP).

Post-hoc power analysis using the G*Power 3 software (Faul, Erdfelder, Buchner, & Lang, 2009) indicated that the present sample size of 69 participants provided a power of 78% to detect interaction effects at the small to medium level with three groups and two repeated measures.

Procedure and Tasks

All tasks were administered in group settings, starting with the arithmetic task.

Arithmetic

To assess children's arithmetic skills, the TTR (Tempo Test Rekenen; De Vos, 1992) was administered. It is a paper-and-pencil timed standardized arithmetic test that consists of 200 arithmetic number fact problems. All items are presented across five DIN A4 sheets: 1 sheet for addition, subtraction, multiplication, division, and a mixture of these operations, respectively. On each sheet, the problems increase in difficulty. Children need to solve as many items as possible within 1 min per sheet (i.e., 5 min in total). In line with the results of previous studies (Kim & Opfer, 2017; Link, Nuerk, et al.,

2014), only the addition and subtraction problems were evaluated for the present study. Children received one point for every correctly solved item (i.e., maximum score = 40 per operation).

NLE

Stimuli and design — The study consisted of a bounded and unbounded version of the NLE task, requiring participants to indicate the correct position of a presented target number on a number line. All number lines had a constant length of 20 cm and were presented on paper in landscape format with one item per DIN A4 sheet.

In the bounded task, number lines were labelled below the origin and endpoint with the numbers 0 and 100, respectively. Target numbers [2, 3, 6, 7, 9, 11, 13, 17, 18, 27, 35, 47, 53, 64, 75, 82, 95, 99] were placed above the origin with one target number per number line. Children were instructed that they are presented with a number line that only has an origin and an endpoint. The task was then explained (in Luxembourgish) as follows: “Look at the number above the number line – where do you think this number goes between 0 and 100. Please mark your estimate on the number line.” 50 was used as practice trial.

In the unbounded task, number lines were only labelled below the origin with 0 but did not comprise a clearly defined endpoint. A unit indicating the distance between 0 and 1 was depicted below the origin. Target numbers were presented above the origin. The numerical length of the unbounded number line was 29. However, only items up to 20 [2, 3, 4, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19] were used to keep sufficient space between the largest target number and the physical endpoint. Children were instructed that there is no end to the number line but that they can see how long the distance from 0 to 1 is. 10 was used as practice trial.

All children first completed the unbounded task followed by the bounded one. This was to avoid that the endpoint ‘100’ given in the latter task biased estimations in the former. We particularly wanted to prevent the children from mistakenly assuming that unbounded number lines might also cover the range from 0 to 100. Participants were not informed about the number range covered by the unbounded task prior to completing it. In other words, the task that explicitly defined a number range (i.e., the bounded task) had to be administered last to avoid participants from potentially building up expectations about the number range used in the unbounded task. This procedure has also been previously used in studies assessing unbounded and bounded NLEs using a within-subject design (e.g., Cohen & Sarnecka, 2014; Link, Huber, et al., 2014; Link, Nuerk, et al., 2014; Reinert et al., 2017; Reinert et al., 2019). No feedback was provided to the children regarding the correctness of their estimates for either practise or experimental trials in both tasks.

Analysis — *Outlier Exclusion.* Individual estimates that differed more than ± 2.5 SDs from the respective grade’s mean estimate were excluded prior to data analyses (cf., Link, Nuerk, et al., 2014; see also Cohen & Sarnecka, 2014). Overall, this resulted in the exclusion of 1.64% and 1.93% of estimates in the unbounded and bounded task respectively. There was no effect of task or grade and no interaction on the percentage of excluded items. Importantly, this trimming procedure did not change any of the main findings.

Percent Absolute Error. The percent absolute error (PAE = $\text{abs}(\text{estimate} - \text{target number})/\text{number scale} \times 100$) was averaged across all target numbers separately for each individual in each task to index overall task difficulty in every child.

We also calculated mean PAEs as a function of target number in both the unbounded and bounded tasks across all children separately for each grade.

Model Fittings. We assessed the goodness of fit of several models previously shown to mathematically reflect children’s NLEs. Models were estimated using conventional ordinary least square regression in R.

We distinguished between models that either index direct estimation and dead-reckoning or reflect calculation strategies, such as proportional judgement. Direct estimation is indicated by the superior fits of either the mixed log-linear model (MLLM) or the scalloped power models (SPMs), with the latter also reflecting some extent of dead-reckoning depending on the variant of the model. Conversely, proportional judgment can be revealed by the superior fits of the cyclic power models (CPMs), including the subtraction bias cyclic model (SBCM).

The MLLM predicts estimates as a weighted sum of logarithmic and linear transforms of the target number (Anobile et al., 2012; Cicchini et al., 2014; Kim & Opfer, 2017; Opfer et al., 2016). It reflects the assumption that logarithmic and

linear representations co-exist to some extent in the same individual, since logarithmic-to-linear shifts occur at different times in development for different number ranges. The MLLM is defined as:

$$y = a((1 - \lambda)x + \lambda \frac{U}{\ln(U)} \ln(x))$$

where y indicates the estimate of a target number x on a $0-U$ number line. a denotes a scaling parameter and λ is a logarithmicity index that measures the degrees of logarithmic compression in estimates. If estimation is perfectly linear, λ converges to 0, whereas its value approaches 1 as estimation shows more logarithmic compression.

The SPMs reflect number estimation via dead reckoning, whereby participants firstly estimate a unit and then estimate the position of the following unit based on their current position. Direct estimation is indexed by a single-scalloped power model (1SPM), which is identical to Stevens' power law without a scaling factor:

$$y = x^\beta$$

Dual- (2SPM) and multi-scalloped power models (multi-SPM) reflect the estimation of a particular working window of numbers (e.g., 5) before using multiples of this working window to estimate the position of higher target numbers. The 2SPM allows to identify participants, estimating the working window once before positioning their estimates:

$$y = x^\beta, \text{ if } x < d$$

$$y = d^\beta + (x - d)^\beta, \text{ if } x \geq d$$

while the multi-SPM indexes multiple applications of the working window:

$$y = \text{truncate}\left(\frac{x}{d}, 0\right) \cdot d^\beta + (x \text{ modulo } d)^\beta$$

In these models, x is the target number, β is the characteristic exponent, describing the numerical bias and the shape of the power function, and d is the size of the working window.

Finally, CPMs index calculation strategies, such as probably proportional judgment. These models suggest that participants use at least two reference points (i.e., the origin and the endpoint) to guide their estimations. Number placement thus occurs via estimating a target number's distance from the lower and upper bounds of the number line until consistency of the two distances. While the one-cycle power model (1CPM) reflects the use of two reference points (i.e., the origin and the endpoint):

$$y = U \frac{x^\beta}{x^\beta + (U - x)^\beta}$$

the two-cycle power model (2CPM) indexes the reliance on an additional central reference point:

$$y = \frac{U}{2} \cdot \frac{x^\beta}{x^\beta + (\frac{U}{2} - x)^\beta}, \text{ if } x \leq \frac{U}{2}$$

$$y = \frac{U}{2} + \frac{U}{2} \cdot \frac{(x - \frac{U}{2})^\beta}{(x - \frac{U}{2})^\beta + (U - x)^\beta}, \text{ if } x > \frac{U}{2}$$

1CPM and 2CPM were fitted with one free parameter, the exponent β describing the numerical bias and the shape of the power function. U represents the upper bound of the number line.

In addition to these models, Cohen and Sarnecka (2014) developed the subtraction bias cyclic model (SBCM) to formally model poor mensuration. To scale the length of a number line to a proportion by estimating the distances from

the origin and the endpoint of the line, participants need to subtract the target number from the value of the upper bound. To account for difficulties with this subtract-and-compare process, the 1CPM was modified to incorporate an exponent s , capturing this potential bias in subtraction:

$$y = U \frac{x^\beta}{x^\beta + ((U - x)^s)^\beta}$$

Importantly, CPMs require definition of an upper bound U , which can be easily specified in the bounded task. However, for the unbounded task, such an upper bound does not exist per se. In theory, participants might have used the end of the physical line as an upper bound (i.e., 29) or the largest target number (i.e., 19). Since these strategies might vary between participants, a fixed upper bound for testing these models in the unbounded task cannot be used. The upper bound was therefore estimated by the fitting procedure for unbounded NLEs. The range of the parameter accounting for the upper bound was constraint to lie between 19 and 29. Unbounded estimates were only fitted with the modified version of the SBCM and 1CPM, but not the more complex 2CPM, indexing the reliance on an additional central reference point.

In both the SPMs and CPMs, accurate responding is indicated by $\beta = 1$, while $\beta < 1$ indicates a negatively accelerating bias (i.e., logarithmic), and $\beta > 1$ indicates a positively accelerating bias (i.e., exponential). β was set to be greater than 0.

The models were fitted both to individual estimates and to median estimates separately for each grade. Models were compared in terms of goodness of fit by calculating AICc¹ (Akaike information criterion with a correction for finite sample sizes) values (e.g., Burnham, Anderson, & Huyvaert, 2011), with lower AICc values indexing superior model fits (see e.g., Barth et al., 2016; Gross et al., 2018; Kim & Opfer, 2017; Link, Huber, et al., 2014; Luwel et al., 2018; Möhring et al., 2018; Reinert et al., 2017; Sasanguie et al., 2016; Slusser et al., 2013, for a comparable use of AIC). AICc is a penalized-likelihood criterion, meaning that it penalizes for the number of parameters in a model. In contrast to other measures (e.g., R^2), it thus considers both goodness of fit and model complexity in terms of the number of parameters (Burnham & Anderson, 2004).

Results

Unbounded and Bounded NLEs

Estimation Errors

Firstly, we determined whether estimation errors depended on grade and/or the version of the number line task. We therefore conducted a linear mixed effect model using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the R environment (R Development Core Team, 2007). Task and grade with interaction term were entered as fixed effects into the model. As random effect, we included intercepts for participants. We then compared this model to reduced models without the effects or interaction in question using chi-squares tests on the log-likelihood values. The full model fitted better than the reduced model without the interaction term, $\chi^2(1) = 8.53$, $p = .003$. The full model was also significantly better than the reduced models excluding either grade, $\chi^2(1) = 6.10$, $p = .01$, or task, $\chi^2(1) = 4.26$, $p = .04$. Performances were better on the bounded (PAE = 7.57, $SD = 3.43$) than the unbounded task (PAE = 13.02, $SD = 5.18$). Moreover, performances increased with grade, but mainly on the bounded task (2nd grade: PAE = 9.25, $SD = 3.94$, 3rd grade: PAE = 8.45, $SD = 3.25$, 4th grade: PAE = 5.97, $SD = 2.43$) compared to the unbounded task (2nd grade: PAE = 11.12, $SD = 4.33$, 3rd grade: PAE = 14.96, $SD = 4.97$, 4th grade: PAE = 13.12, $SD = 5.51$). In addition, there was no correlation between unbounded and bounded mean PAEs ($r = -.04$, $p = .74$) or SD s of mean PAEs ($r = -.004$, $p = .97$). This did not change when controlling for grade (mean PAE: $r = .02$, $p = .89$, SD of mean PAE: $r = .04$, $p = .75$). Mean PAEs did, however, relate to SD s of mean PAEs in both the unbounded ($r = .81$, $p < .001$) and bounded tasks ($r = .86$, $p < .001$).

1) The same results were obtained when using the Bayesian information criterion (BIC) instead of AICc as a measure of goodness of fit (see also Dackermann et al., 2015; Opfer et al., 2016).

Next, we correlated unbounded and bounded PAEs with the size of target numbers separately for each grade. Unbounded PAEs significantly increased with increasing target number in every grade (2nd grade: $r = .93$, $p < .001$, 3rd grade: $r = .93$, $p < .001$, 4th grade: $r = .93$, $p < .001$). Conversely, no relations between PAEs and the size of target numbers were observed in the bounded task, except for third grade, where estimation errors were smaller for larger target numbers ($r = -.7$, $p < .001$). Visual inspection of the plots indicated a pattern of bounded PAE distribution reflecting fewer errors in and around to-be-expected reference points, including the origin and endpoint (reference points 2, 3, 47, 53, 95, 99: mean PAE = 5.48, $SD = 2.28$ vs. no reference points 7, 11, 17, 27, 64, 75, 82: mean PAE = 8.00, $SD = 3.76$), which is characteristic of proportional judgement (Ashcraft & Moore, 2012). This thus strengthens the assumption that different constructs underlie unbounded and bounded NLEs.

Model Fits and Parameters

To further confirm the use of different strategies on unbounded and bounded tasks, we fitted a series of models to children's unbounded and bounded NLEs. Since relying only on mean or median estimates across individuals from each grade for each task might obscure individual differences in estimation patterns and trajectories (see also Sasanguie, Verschaffel, Reynvoet, & Luwel, 2016), all models were fit to individual estimates in a first step. However, for the sake of completeness and to be in accordance with some previous studies (Barth & Paladino, 2011; Cohen & Sarnecka, 2014; Slusser & Barth, 2017; Slusser et al., 2013), model fits were also assessed at the grade level by considering unbounded and bounded grade median estimates.

First, children were grouped based on the model that best described their individual estimates. Table 1 presents the percentage of children best fit by the different models separately for each task and grade. In the unbounded task, the estimates of approximately 50% of the children were best fit by the 1SPM in each grade, while the MLLM explained NLEs in about one third of all the children in every grade. Overall, unbounded NLEs of less than 10% of the children were best described by the multi-SPM or any of the CPMs. Conversely, in the bounded task, the CPMs provided the best fit for the estimates of about 90% of the children in each grade. While the bounded estimates of most of the second and third graders were best explained by the SBCM (i.e., 40%), the estimates of fourth graders were described to an equivalent extent by either the SBCM, 1CPM or 2CPM. The bounded estimates of only 10% of the children in every grade were best described by the MLLM, while the SPMs provided the best fit for the bounded estimates of almost none of the children².

To determine whether NLEs in terms of PAEs depended on the strategy used to position target numbers on either the unbounded or bounded number line (i.e., best-fit model), we performed two one-way ANOVAs on either unbounded or bounded PAEs including best-fit unbounded or bounded model respectively as between-subject factor. For these analyses, individuals whose unbounded NLEs were not described by either the MLLM or 1SPM were pooled to avoid very small sample sizes. For the same reason, those participants whose bounded NLEs were not fit by a variant of the CPMs were combined. Analysis revealed a main effect of best-fit model for bounded PAEs, in that children using three reference points outperformed children whose estimates were best described by either the MLLM or 1SPM (2CPM: PAE = 5.47, MLLM-1SPM: PAE = 9.32, $F(3, 64) = 4.33$, $p = .008$, $\eta_p^2 = .17$), even when controlling for grade. No other differences were observed for bounded NLEs. For unbounded PAEs, there was also a main effect of model, such that individuals whose estimates were described by the 1SPM performed better than children whose estimates were fit by the MLLM (MLLM: PAE = 14.71, 1SPM: PAE = 11.44, $F(2, 65) = 3.58$, $p = .034$, $\eta_p^2 = .10$).

2) It could be argued that the better fits of the CPMs compared to the MLLM in the bounded task might be explained by a decline in attentional processes on this task rather than by its boundedness. Namely, task order was fixed in the present study with the bounded task always being administered last and attention was previously shown to generally increase the linearity of NLEs (see e.g., Anobile et al., 2012). Nonetheless, we believe that this is unlikely the case since some children's NLEs were best-fit by a linear function in the bounded but not in the unbounded task, although the former task was administered last. Together with the aforementioned findings of the lack of correlation between unbounded and bounded PAEs and the qualitative differences in the effect of grade on the latter, we assume that any differences between unbounded and bounded NLEs (in terms of PAE and/or best-fit model) were unlikely confounded by task order and attentional processes.

Table 1*The Percentage of Children Best Fit by Each Model for Each Task and Grade*

Grade	MLLM	1SPM	2SPM	Multi-SPM	SBCM	1CPM	2CPM
Unbounded Task							
2 nd	30	55	0	5	5	5	–
3 rd	33.33	50	0	5.56	11.11	0	–
4 th	38.71	48.39	0	6.44	3.23	3.23	–
All	34.78	50.72	0	5.80	5.80	2.90	
Bounded Task							
2 nd	10	0	0	0	40	25	25
3 rd	11.11	0	0	0	44.44	27.78	16.67
4 th	9.68	6.45	0	0	25.81	29.03	29.03
All	10.14	2.90	0	0	34.78	27.54	24.64

Apart from this, we also considered the average goodness of fit, as indexed by AICc, across all participants per grade for each model used to fit either unbounded or bounded NLEs. Mean unbounded and bounded AICc are displayed in Table 2. According to Burnham and Anderson (2002), models having a ΔAICc within 0–2 of the best model have substantial support and should be taken into consideration when making inferences, while models with a ΔAICc within 4–7 have considerably less support and models with a $\Delta\text{AICc} > 10$ have essentially no support. In the unbounded task, the 1SPM ($\text{AICc} = 26.65$) provided the best fit across all individuals, but the MLLM and the remaining SPMs could also substantially support the data (i.e., $\Delta\text{AICc} < 2$, see Table 2). The worst fits were provided by the CPMs, notably the modified version of the 1CPM including an additional free parameter that represented the “upper bound” of the unbounded number line. The numerical bias captured by the exponent β of the 1SPM did not depend on grade, $F(2, 66) = 2.83$, $p = .07$, $\eta_p^2 = .08$, and was significantly positively accelerating across all grades, $\beta = 1.09$, $t(68) = 11.41$, $p < .001$. Similarly, grade had no effect on λ , which approached 0, $\lambda = 0.098$, $F(2, 66) = 0.21$, $p = .81$, $\eta_p^2 = .01$. These findings thus provide no evidence for logarithmic-like estimation patterns on unbounded number lines in the current sample. In the bounded task, the SBCM provided the best fit across all participants ($\text{AICc} = 67.04$) as well as in each grade. According to Burnham and Anderson (2002), the bounded NLEs of fourth graders could also be explained by the 1CPM (i.e., $\Delta\text{AICc} < 2$), while those of second and third graders were only best described by the SBCM (see Table 2). The 2CPM only provided moderate support for the children’s estimates even in fourth graders, where the estimates of almost one third of the children were best fit by this model (see Table 2). Overall, the outcome suggests that children generally rely on two reference points (i.e., the origin and endpoint) with some bias in subtraction, when estimating the position of target numbers on bounded number lines. Interestingly, the exponent s of the SBCM, capturing this subtraction bias, did not differ between grades, $F(2, 66) = 2.18$, $p = .12$, $\eta_p^2 = .06$, and was relatively close to 1 in all grades (2nd grade: $s = 0.96$, 3rd grade: $s = 0.96$, 4th grade: $s = 1.00$). Conversely, the exponent β depended on grade for both the SBCM, $F(2, 66) = 8.86$, $p < .001$, $\eta_p^2 = .21$, and 1CPM, $F(2, 66) = 8.96$, $p < .001$, $\eta_p^2 = .21$. It was negatively accelerating in each grade, but approached 1 as children got older, reflecting progressively linear estimations (SBCM: 2nd grade = 0.69, 3rd grade = 0.71, 4th grade = 0.89; 1CPM: 2nd grade = 0.66, 3rd grade = 0.68, 4th grade = 0.89). Importantly, unbounded and bounded model parameters did not correlate (see Table 3), further indicating the reliance on different cognitive constructs when estimating the position of target numbers and either unbounded or bounded number lines.

Table 2*Mean AICc of Each Model for Each Task and Grade*

Grade	MLLM	1SPM	2SPM	Multi-SPM	SBCM	1CPM	2CPM
Unbounded Task							
2 nd	28.20	27.43	29.75	29.49	31.18	42.12	–
3 rd	28.44	28.77	30.33	29.98	34.57	51.21	–
4 th	23.66	23.75	25.99	24.96	29.33	46.87	–
All	26.77	26.65	28.69	28.15	31.70	46.74	
$\Delta AICc$	0.12	0.00	2.04	1.50	5.05	20.09	
Bounded Task							
2 nd	72.85	85.60	88.48	88.62	70.67	74.01	79.28
3 rd	71.05	83.84	86.71	86.70	67.73	70.39	74.67
4 th	67.67	71.69	74.55	74.47	62.71	64.46	67.10
All	70.53	80.38	83.25	83.27	67.04	69.62	73.68
$\Delta AICc$	3.49	13.34	16.21	16.23	0.00	2.58	6.64

Note. $\Delta AICc$ represents the difference to the best model across all grades separately for each task. The best model is indicated by $\Delta AICc = 0$.

Table 3*Correlations Between Unbounded and Bounded Model Parameters*

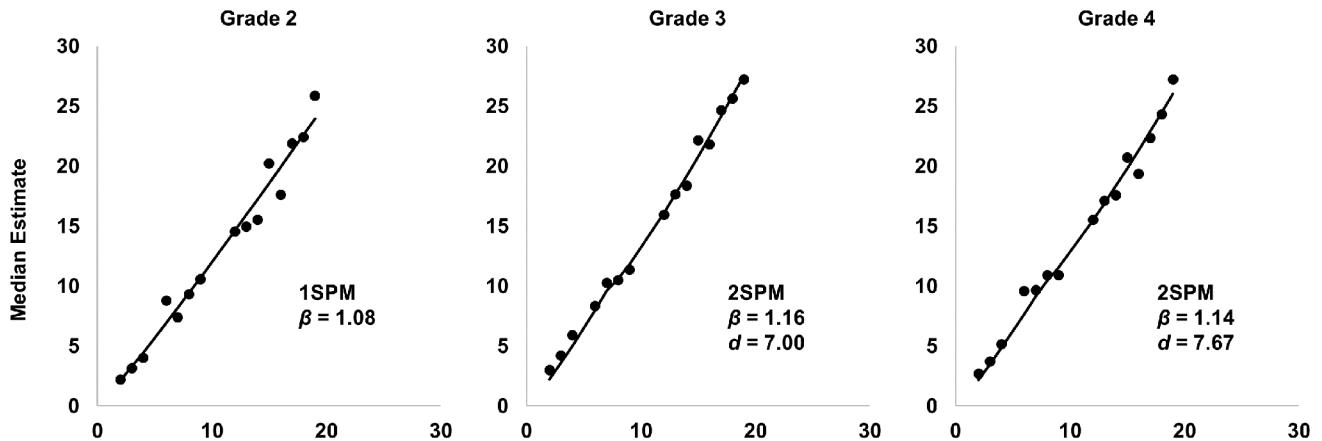
Parameter	Correlation Coefficient
$MLLM_{\lambda}$	-.03
$1SPM_{\beta}$.11
$2SPM_{\beta}$.12
$Multi-SPM_{\beta}$.07
SBCMs	.14
$SBCM_{\beta}$.19
$1CPM_{\beta}$.14

Alongside these analyses at the individual level, we also fitted the different models to the children's median estimates separately for each task and grade (see Figure 1). Median estimates in the unbounded task were best described by the 1SPM in second graders and the 2SPM in third and fourth graders. In the bounded task, the SBCM, 1CPM and 2CPM provided the best fits for median estimations of second, third, and fourth graders respectively. However, since median estimates are computed across participants, possibly differing from each other in terms of the best-fit model, one should be cautious when interpreting the outcomes (Firebaugh, 2015).

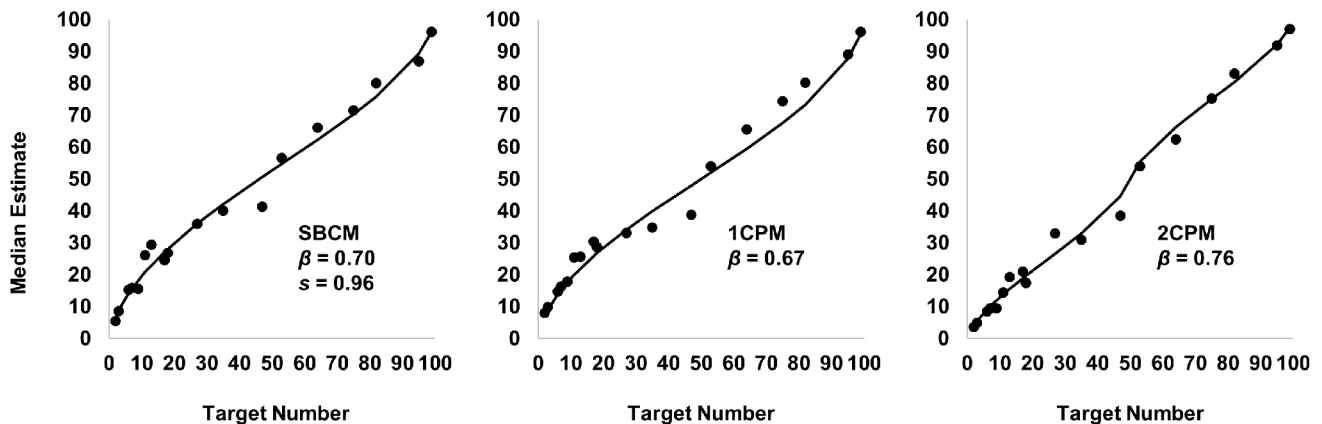
Figure 1

Median Estimates by Target Number in the Unbounded and Bounded Tasks for Each Grade

Unbounded



Bounded



Note. Solid lines represent best-fit models. Unbounded median estimates were best fit by the 1SPM in second graders and the 2SPM in third and fourth graders. Bounded median estimates were best fit by the SBCM, 1CPM, and 2CPM in second, third, and fourth graders, respectively.

Arithmetic Skills

To determine whether arithmetic performances depended on grade and/or task (addition vs. subtraction), we also conducted a linear mixed effect model. Task and grade with interaction term were entered as fixed effects, while we included intercepts for participants as random effect. Model comparisons were done using chi-squares tests on the log-likelihood values. Since the full model did not provide a better fit than the reduced model without the interaction term, $\chi^2(1) = 2.53$, $p = .11$, we proceeded with contrasting the model without interaction term to models that were further reduced by excluding the remaining effects in question. Model fits were significantly worse when excluding either grade, $\chi^2(1) = 30.61$, $p < .001$, or task, $\chi^2(1) = 43.47$, $p < .001$. Performances were weaker in the subtraction (17.67, $SD = 4.52$) than the addition task (21.17, $SD = 5.14$) and in younger children (2nd grade = 31.70, 3rd grade = 37.56, 4th grade = 44.19). Performances were significantly correlated in both tasks ($r = .7$, $p < .001$).

Relations Between NLEs and Arithmetic Skills

Better addition and subtraction skills were related to significantly fewer and less variable estimation errors, as indexed by individual mean PAEs and SDs of mean PAEs, in the bounded but not the unbounded task (see Table 4, above the

diagonal). Controlling for grade did, however, affect the relation between bounded NLEs and subtraction skills (see Table 4, below the diagonal).

Table 4

Correlations Between Unbounded and Bounded NLEs and Arithmetic Skills

Measure	1	2	3	4	5	6
1. Addition	–	.702***	.001	-.008	-.422***	-.437***
2. Subtraction	.578***	–	.144	.138	-.368**	-.384**
3. Unbounded PAE (<i>M</i>)	-.099	.087	–	.812***	-.041	-.107
4. Unbounded PAE (<i>SD</i>)	-.069	.113	.812***	–	.006	-.004
5. Bounded PAE (<i>M</i>)	-.240*	-.198	.017	.043	–	.856***
6. Bounded PAE (<i>SD</i>)	-.219 [†]	-.184	-.048	.039	.823***	–

Note. Partial correlations controlling for grade are displayed below the diagonal.

[†] $p < .07$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Since most of the children relied on calculation strategies, such as probably proportional reasoning, when placing target numbers on bounded number lines, we additionally assessed correlations with arithmetic skills when focussing only on those individuals whose bounded NLEs were best described by one of the variants of the CPMs ($n = 60$). Correlations between arithmetic skills and bounded mean PAEs as well as *SDs* of mean PAEs remained stable compared to including all children (Fisher's z for comparison of correlations based on independent samples: addition: mean PAE: $z = 0.14$, $p = .89$, *SD* of mean PAE: $z = -0.05$, $p = .96$, subtraction: mean PAE: $z = 0.41$, $p = .68$, *SD* of mean PAE: $z = 0.40$, $p = .69$, see Table 5, above the diagonal). This suggests that relations with arithmetic skills were not driven by individuals that did not apply a calculation strategy, such as proportional judgement ($n = 9$). Moreover, controlling for grade had no effect on the significance of the relations between arithmetic skills and bounded NLEs, when including only those children likely using proportional reasoning (see Table 5, below the diagonal).

Table 5

Correlations Between Bounded NLEs and Arithmetic Skills in Children Whose Performances Are Best Described by the CPMs ($n = 60$)

Measure	1	2	3	4
1. Addition	–	.735***	-.442***	-.429**
2. Subtraction	.611***	–	-.431**	-.444***
3. Bounded PAE (<i>M</i>)	-.319*	-.315*	–	.825***
4. Bounded PAE (<i>SD</i>)	-.249 [†]	-.289*	.800***	–

Note. Partial correlations controlling for grade are displayed below the diagonal.

[†] $p < .06$. * $p < .05$. ** $p < .01$. *** $p < .001$.

To further assess the importance of calculation strategies, we determined whether the relation between arithmetic skills and bounded NLEs was conditional upon strategy use by performing moderation analysis using Hayes' PROCESS macro for SPSS. In two separate analyses, we assessed the effect of bounded mean PAEs on addition and subtraction skills respectively, including best-fit model as moderator. Those children whose estimates were best fit by either the MLLM or the 1SPM were categorized as not applying any calculation strategy and compared to those individuals classified as either SBCM, 1CPM or 2CPM. This resulted in a multi-categorical moderator of four levels. In this model, moderation is depicted by the significant effect of the interaction term between bounded mean PAEs and best-fit model on addition and/or subtraction skills, while controlling for the effects of the factors included in the interaction term. A bootstrapping approach with 5.000 bootstrap samples was used. Significance was determined at 95% bias-corrected confidence intervals. To avoid multicollinearity issues, all variables were mean centred prior to analyses.

While the interaction between bounded mean PAEs and best-fit model did not account for a significant proportion of the variance in addition skills ($\Delta R^2 = .06$, $p = .13$), a tendency was observed for subtraction skills ($\Delta R^2 = .08$, $p = .08$). When examining the conditional effect at the four different levels of the moderator, no significant relation between bounded mean PAEs and subtraction skills was observed in children that did not apply any calculation strategy, $b = -0.03$, $t(61) = -0.10$, $p = .92$. Conversely, as expected, fewer bounded estimation errors were associated with significantly better subtraction skills in children whose estimates were best described by either the SBCM, $b = -0.88$, $t(61) = -4.00$, $p < .001$, the 1CPM, $b = -0.93$, $t(61) = -2.57$, $p = .01$, or the 2CPM, $b = -1.06$, $t(61) = -1.99$, $p = .05$.

We also assessed correlations between addition and subtraction skills and best-fit model parameters (see Table 6). Model parameters were adjusted as such: absolute values of $s-1$ and $\beta-1$. This allowed us to directly determine whether larger deviations from 1 were associated with poorer arithmetic skills. Parameters of the best-fit models in the unbounded task (i.e., λ of MLLM and β of 1SPM) did not correlate with either addition or subtraction skills. Conversely, significant correlations were observed between arithmetic skills and parameters of the models best describing bounded NLEs (i.e., s of SBCM and β s of SBCM, 1CPM and 2CPM). More concretely, smaller biases (i.e., parameter values closer to 1) were associated with better arithmetic skills. Interestingly, parameter s of SBCM, capturing the extent of subtraction bias, was related to addition but not subtraction skills.

Table 6

Correlations Between the Parameters of the Best-Fit Models for Each Task and Arithmetic Skills

Parameter	Addition	Subtraction
Unbounded MLLM (λ)	.088	-.072
Unbounded 1SPM (β)	.012	.179
Bounded SBCM (s)	-.307*	-.170
Bounded SBCM (β)	-.274*	-.272*
Bounded 1CPM (β)	-.263*	-.220 [†]
Bounded 2CPM (β)	-.330**	-.308**

[†] $p < .07$. * $p < .05$. ** $p < .01$.

Finally, we performed a linear mixed model to determine the effects of best-fit CPM (SBCM vs. 1CPM vs. 2CPM) and/or task (addition vs. subtraction) on arithmetic skills to assess whether the latter performances were affected by the specific type of proportional judgement strategy applied by the children. Best-fit CPM and task with interaction term were entered as fixed effects and intercepts for participants as random effect. Model comparisons were done using chi-squares tests on the log-likelihood values. Since the full model did not provide a better fit than the reduced model without the interaction term, $\chi^2(1) = 0.54$, $p = .46$, we proceeded with contrasting the model without interaction term to models that were further reduced by excluding the remaining effects in question. Model fits were significantly worse when excluding task, $\chi^2(1) = 40.24$, $p < .001$, but not best-fit CPM, $\chi^2(1) = 1.19$, $p = .28$. This thus indicates no differences in overall arithmetic skills depending on the specific calculation strategy used, with additions always being easier than subtractions. In sum, our findings suggest that calculation strategies, such as probably proportional reasoning, underlie the relation between NLEs and arithmetics. However, using a more sophisticated strategy with three instead of two reference points does not positively affect arithmetic skills or vice-versa better arithmetic skills do not necessarily entail the reliance on more reference points when estimating the position of target numbers on bounded number lines. Conversely, how well you apply a certain calculation strategy depends on arithmetic abilities, considering the relations of the different model parameters (i.e., β s) with addition and subtraction skills.

Discussion

In this paper, we aimed to shed further light on the cognitive mechanisms underlying the commonly observed relation between NLEs and arithmetics in older elementary school children. Below we will first discuss the cognitive constructs

indexed by the two different NLE tasks and then consider how those mechanisms could explain the differential relations of unbounded and bounded NLEs to arithmetic performances.

Unbounded and Bounded NLEs Likely Index Different Cognitive Constructs

Unbounded and bounded NLEs were unrelated in terms of both estimation errors and model parameters. Moreover, only bounded NLEs improved with grade and were overall better than unbounded performances. In addition, while unbounded NLEs were best described by models reflecting direct estimation, functions supposedly capturing proportional judgment provided better fits for bounded NLEs. The two tasks thus probably elicit different estimation strategies with only the unbounded version providing an accurate measure of children's numerical magnitude representations in the current sample.

More concretely, unbounded NLEs of about 85% of the children were best described by either the MLLM or the 1SPM, reflecting direct estimation. In addition, the goodness of fits (as indexed by AICc) provided by these models were better than those of the CPMs, probably indexing proportional reasoning strategies. Unbounded estimation errors also significantly increased as a function of target number. This pattern reflects a signature of the approximate number system (e.g., Cantlon, Cordes, Libertus, & Brannon, 2009; Cohen & Blanc-Goldhammer, 2011; Gallistel & Gelman, 2000) and is not characteristic of proportional judgment, indicated by an M-shaped distribution of error variability (e.g., Link, Huber, et al., 2014; Reinert et al., 2019). Our findings thus suggest that the unbounded task directly assesses the representation of numerical magnitudes. This agrees with previous studies, reporting the superior fits of models directly tapping into number representation (Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Jung et al., 2020; Kim & Opfer, 2017; Link, Huber, et al., 2014; Reinert et al., 2017). Interestingly, there were no grade-related differences in the degree of logarithmic compression in estimations, as indexed by λ of the MLLM, or the numerical bias in estimations, as reflected by β of 1SPM, suggesting that numerical magnitude representations remained fairly stable at that developmental stage. Moreover, λ was close to 0 and β was significantly positively accelerating across all grades. These findings thus provide no evidence for any logarithmic compression in unbounded estimations, even in the youngest children of the current sample, and thereby confirm the assumption that NLEs are no longer characterized by a logarithmic-like pattern in the current age group using a familiar number range (Booth & Siegler, 2006).

Conversely, bounded NLEs of the majority of the children were best fit by the CPMs, commonly reported to index proportional judgement (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Jung et al., 2020; Rouder & Geary, 2014; Slusser et al., 2013). While the SBCM provided the best fit for estimates of most of the second and third graders, the estimates of the majority of fourth graders were described to an equivalent extent by either the SBCM, 1CPM or 2CPM. This pattern was confirmed when considering the average goodness of fits of all the different models per grade. Based on AICc, both the SBCM and the 1CPM can provide substantial support in favor of bounded NLEs (e.g., Burnham & Anderson, 2002). Model fits thus suggest that in the bounded task, the majority of the children relied on two reference points (i.e., the origin and endpoint) and used subtraction or division to scale target numbers to line length, with some children (mainly the younger) still lacking the mathematical skills to successfully do so. The use of calculation strategies as opposed to direct estimation in the bounded task is also corroborated by the absence of a significant relation between bounded estimation errors and target numbers. Moreover, fewer errors were made at to-be-expected reference points (i.e., the origin and endpoint), further confirming the reliance on proportional judgement in the bounded task. The application of calculation strategies might then also explain grade-related changes in bounded but not unbounded NLEs, considering that such strategies are fostered in school.

Proportional Reasoning Likely Explains the Relation Between NLEs and Arithmetic Skills

Bounded NLEs, likely indexing proportional reasoning, related to addition and subtraction skills. Importantly, these relations remained significant when controlling for grade in children whose estimated were best described by the CPMs. This generally agrees with previous findings, consistently reporting strong relations between bounded NLEs and arithmetics (e.g., Ashcraft & Moore, 2012; Booth & Siegler, 2006; Fazio et al., 2014; Geary, 2011; Jung et al., 2020; Link, Nuerk, et al., 2014; Sasanguie et al., 2013; Siegler & Booth, 2004; see Schneider et al., 2018, for a meta-analysis; see Siegler, 2016, for a review). Conversely, no relation was observed for unbounded NLEs, reflecting direct estimation

in the majority of children. This is also in line with previous observations, indicating that correlations between NLEs and mathematical skills were moderated by the variant of the number line task (Schneider et al., 2018), with positive correlations being observed only for bounded but not unbounded NLEs in older elementary school children (Link, Nuerk, et al., 2014) and adolescents (Jung et al., 2020). Altogether, these outcomes suggest that the commonly observed association between NLEs and mathematical competences is likely explained by calculation strategies rather than the MNL, at least in second to fourth graders. This can be further supported by the moderation analysis, which revealed a tendency for a significant interaction of bounded NLEs and best-fit model on subtraction skills. More concretely, relations between bounded estimation errors and subtraction skills were only observed in children whose estimates were best described by the CPMs, likely capturing proportional judgement, but not the MLLM or 1SPM, indexing direct estimation. It should, however, be noted that the present moderation analysis was mostly exploratory. Namely, groups based on best-fit model were fairly unequal in size, since bounded NLEs of only 10% of the children were best described by models other than the CPMs. Nonetheless, it strengthens the assumption that calculation strategies account for the correlation between NLEs and arithmetics in older elementary school children.

Practical Implications

It is important to also comment on the more practical implications of the present outcomes. What does the absence of a relation between arithmetics and unbounded NLEs, probably providing a purer measure of numerical magnitude representations, tell us about the importance of the latter for arithmetic development? The present findings suggest that the scaling of numerical magnitudes on the MNL does not relate to arithmetics in second to fourth graders. This agrees with studies reporting the absence of a relation between the SNARC effect, another important marker of the MNL, and arithmetic skills in older children attending fourth (Georges et al., 2017) or fifth grade (Schneider et al., 2009) and even adults (Cipora & Nuerk, 2013; but see Hoffmann, Mussolin, Martin, & Schiltz, 2014). In contrast, younger children might rely on the MNL when performing arithmetics. Accordingly, Kim and Opfer (2017) reported that not only bounded but also unbounded NLEs, both best described by the MLLM indexing numerical magnitude representations, significantly correlated with addition and subtraction skills in younger children attending preschool, first or second grade when assessed on relatively unfamiliar number lines. Moreover, some previous SNARC studies indicated strong relations between the SNARC effect and arithmetic skills in preschool (Hoffmann et al., 2013) and up to third grade (Georges et al., 2017). In contrast to the latter SNARC study, we did, however, not find a relation between unbounded NLEs and addition or subtraction skills even in the youngest children attending second grade. This might be explained by the possibility that relations to arithmetics depend on the specific properties of the MNL indexed by the numerical task. The spatial properties of the MNL include both directionality (i.e., the specific directional orientation) and scaling (i.e., the spatial intervals between numerical values, see Aulet & Lourenco, 2018). While the unbounded task likely reflects the scaling of numerical magnitudes, the SNARC task probably rather indexes the directionality of numerical magnitude representations (Cipora, Patro, & Nuerk, 2015). As such, MNL scaling could be more important for arithmetics at the initial stages of mathematical development, while directionality plays an important role also at later stages until none of the properties of the MNL are predictive of math achievement. Once that stage is reached, arithmetic performances might solely depend on calculation strategies, as captured by the bounded task. Overall, the present findings do not question the importance of the MNL, but only suggest that the scaling of numerical magnitude representations might be less important for arithmetic learning in older elementary school children.

Despite the outcome of the present study providing no evidence for the importance of numerical magnitude representations for arithmetic learning, the relation of bounded NLEs to addition and subtraction skills further suggests that bounded number lines are a valuable and robust tool for predicting mathematical competence. Since number line tasks are easily applicable, relatively short, and very cost-effective, they could also be used to assist the diagnosis of mathematical learning difficulties. The present study also further highlights the potential usefulness of training the positioning of target numbers on bounded number lines for arithmetic development.

Limitations and Future Directions

First of all, it should be emphasized that the present outcomes might not be generalizable to different age groups and/or number ranges. We tested older elementary school children to 1) complement previous within-subject designs assessing both model fits and relations to arithmetics in younger children (Kim & Opfer, 2017) or adolescents (Jung et al., 2020) and to 2) determine whether relations between arithmetics and bounded but not unbounded NLEs previously observed in fourth graders (Link, Nuerk, et al., 2014) can actually be explained by the reliance on proportional judgement on bounded as opposed to the MNL on unbounded number lines in such older elementary school children (Link, Huber, et al., 2014). To do so, we focussed on familiar numbers in the range of 0-20 and 0-100 on unbounded and bounded lines respectively (see Link, Huber, et al., 2014 and Link, Nuerk, et al., 2014, for comparable differences in number ranges between bounded and unbounded tasks), since the likelihood of strategy application and the reliable use of reference points was shown to be increased by employing a familiar number range (e.g., Slusser et al., 2013; White & Szűcs, 2012). Furthermore, a familiar number range was also used in the study of Link, Nuerk, et al. (2014), reporting a relation between bounded but not unbounded NLEs and arithmetics in fourth graders.

Since relatively older children (and adults) usually produce seemingly linear estimation patterns on familiar number lines, while negatively accelerating logarithmic-like responses are commonly observed in younger children or with less familiar number ranges, it is probable that different conclusions regarding the constructs underlying bounded and unbounded NLEs as well as their relations to arithmetics might be drawn depending on age and/or number range. Since the shift from seemingly logarithmic to more linear (unbounded and bounded) NLEs with age and experience was suggested to reflect 1) changes in the disposition of numerical magnitude representations (e.g., Siegler & Booth, 2004; Siegler & Opfer, 2003), 2) the maturation of calculation strategies (e.g., the successful use of reference points, see Slusser et al., 2013), 3) the development of better mensuration skills (e.g., Cohen & Sarnecka, 2014) or 4) the reliance on different cognitive constructs (e.g., MNL and calculation strategies in younger and older individuals respectively, Dackermann et al., 2015), all of these factors could potentially differentially influence relations to arithmetics across development.

The idea that different cognitive constructs might underlie NLEs at different developmental stages could be supported by the findings of Yuan and colleagues (2020), showing that performances on the bounded task were related to counting skills and did not correlate with a clustered dot array task that is readily solvable by proportional reasoning skills in 4- to 6-year-olds. This suggests that as opposed to older and more experienced children, bounded NLEs in younger individuals might not index calculation strategies, such as proportional reasoning, but reflect numerical magnitude representations. This would also agree with studies suggesting that strategy application might only develop in older children on familiar number lines (e.g., White & Szűcs, 2012). It could also provide an explanation for the findings of Kim and Opfer (2017), indicating that the negatively accelerating (logarithmic-like) estimation patterns of younger children attending preschool, first or second grade on less familiar bounded number lines were better fitted by the MLLM, indexing the MNL, than by alternative models reflecting calculation strategies. Likewise, Link, Huber, et al. (2014) reported similar results for bounded and unbounded number lines in first graders, suggesting that both tasks might assess the same construct, namely numerical magnitude representations, in these relatively younger children. On the other hand, Cohen and Sarnecka (2014) argued that the negatively accelerating pattern on bounded lines in younger children or with higher number ranges might still describe the application of a calculation strategy (rather than the MNL), yet result from relatively poorer mensuration skills. Likewise, Slusser et al. (2013) reported that younger children's bounded NLEs were best described by the proportion judgment account and that their logarithmic-like estimation pattern reflected the inability to correctly use the upper reference point. In any case, these findings collectively suggest that bounded NLEs might not necessarily index the same advanced proportional reasoning strategy in younger less experienced children than in older children on familiar number lines.

Apart from the constructs underlying NLEs, also their relations to arithmetics might vary depending on age and/or number range. For instance, Kim and Opfer (2017) reported that not only bounded but also unbounded NLEs significantly correlated with addition and subtraction skills in younger children attending preschool, first or second grade when assessed on relatively unfamiliar number lines³. Interestingly, as already mentioned before, in that study both bounded and unbounded NLEs were best described by the MLLM, indexing numerical magnitude representations. This thus agrees with previous studies highlighting the importance of the MNL for arithmetic learning at earlier

developmental stages (e.g., Hoffmann et al., 2013). λ of the MLLM also significantly differed from zero for both bounded and unbounded NLEs, confirming the negatively accelerating logarithmic-like estimation patterns usually observed at this developmental stage. It is therefore possible that as long as unbounded NLEs are characterized by a logarithmic-like pattern, possibly reflecting logarithmically compressed numerical magnitude representations (Kim & Opfer, 2017), they explain variance in arithmetics, with less compression indicating better performances. Since Kim and Opfer (2017) could draw similar conclusions for bounded than unbounded NLEs at such earlier developmental stages, it is likely that proportional reasoning only underlies the relation between bounded NLEs and arithmetics in older children, as assessed in the current study.

Another important point worth mentioning is that due to the linear-like estimation patterns usually observed in older children on familiar number lines, the present study did not include any additional models, such as the bilinear (also known as decomposed linear or segmented linear) account. This model was suggested to provide good fits for logarithmic-like NLEs and is thereby yet another alternative explanation for the logarithmic-to-linear shift hypothesis (see Ebersbach et al., 2008 or Moeller et al., 2009). More concretely, this account suggests that the logarithmic-like shape of NLEs in younger children or on unfamiliar number lines results from a difference in numerical processing depending on the number interval, with each number segment yielding a different slope. While Ebersbach and colleagues (2008) suggested that the breakpoint at which the response function alters reflects changes in number familiarity, Moeller and colleagues (2009) argued that it might depend on the understanding of the place-value structure of the Arabic number system. Since such an estimation pattern originating from two linear intervals can also be fitted by a logarithmic curve, these authors argued that the underlying numerical magnitude representations might have been erroneously assumed to feature a logarithmic disposition in younger children. Instead, they argue for a decomposed representation of either single- and two-digit numbers or of familiar and unfamiliar numerical magnitudes. This decomposed representation is then eventually integrated into one holistic linear representation with age and experience, explaining mostly linear-like NLEs in older children. Including this account as a potential model for negatively accelerating response patterns in younger children or with less familiar numbers could shed further light on whether the initially logarithmic-like response patterns on unbounded and bounded number lines actually index a logarithmically compressed MNL (e.g., Siegler & Booth, 2004; Siegler & Opfer, 2003), relatively poorer mensuration skills (e.g., Cohen & Sarnecka, 2014), the inability to use reference points (e.g., Slusser et al., 2013), or rather insufficient place-value coding (Moeller et al., 2009).

Finally, apart from using different models with unfamiliar number ranges at earlier developmental stages, future studies should also complement the present findings by considering potential domain-specific and/or domain-general covariates. In the current study, we employed a hybrid design, where we used a modelling procedure to better understand the constructs underlying (bounded and unbounded) NLEs and consequently individual differences in addition and subtraction skills (see e.g., Kim & Opfer, 2017, or Jung et al., 2020, for a similar design). Nonetheless, further assessing the influences of basic numerical skills and/or visuospatial abilities on NLEs and arithmetics as well as their relations might strengthen the current conclusions. Based on the present outcomes, one would for instance envisage a relation between visuospatial abilities, likely required for efficient proportional reasoning, and bounded but not (or to a lesser extent) unbounded NLEs. Visuospatial abilities might then also act as a confounder in the relation between bounded NLEs and arithmetics. Support for this assumption has already been provided by some studies, showing that the relation between bounded NLEs and mathematical performances was fully explained by visuospatial abilities in 10-year-old children (Simms et al., 2016) and adults (Sella et al., 2016). Moreover, since only unbounded but not bounded NLEs were found to provide a reliable index of the MNL in the current sample using a familiar number range, only the former might relate to other measures of symbolic and non-symbolic numerical magnitude representations, such as (non-)symbolic number comparison performances, the distance and/or SNARC effects. Evidence for this idea already comes from the study of Schneider and colleagues (2009), finding no relations between bounded NLEs and the SNARC or distance effect in fifth and sixth graders.

3) Considering that Kim and Opfer (2017) presented participants with unbounded number lines up to 1000, but likely did not provide enough space to validly measure a positively accelerating response function, unbounded NLEs could have been inadvertently biased similarly to bounded NLEs, such that their conclusions based on unbounded lines should be taken with care (see also Cohen & Ray, 2020).

Conclusion

In second to fourth graders, unbounded and bounded NLEs index different cognitive constructs. While unbounded estimates reflect direct estimation, thereby providing an appropriate measure of the scaling of numerical magnitude representations, bounded estimates rather index calculation strategies, such as proportional reasoning. These calculation strategies then likely account for the relation of bounded but not unbounded NLEs to addition and subtraction skills. Although the present findings do not provide any evidence for the involvement of numerical magnitude representations for arithmetic learning, we cannot rule out their importance at earlier developmental stages. Depending on the aim of future studies, the present outcomes suggest measuring estimations on unbounded number lines if one is interested in directly assessing numerical magnitude representations in second to fourth graders. Conversely, if one aims to predict arithmetic skills at this age, one should rather assess estimations on bounded number lines, likely indexing proportional reasoning.

Funding: The current research was supported by the National Research Fund Luxembourg (FNR, www.fnr.lu) under Grant AFR PhD-2013-1/5558196.

Acknowledgments: The authors would like to thank Charlotte Sosson for her help with data collection.

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

References

- Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*, 454-459. <https://doi.org/10.1016/j.cognition.2011.11.006>
- Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*, 246-267. <https://doi.org/10.1016/j.jecp.2011.08.005>
- Aulet, L. S., & Lourenco, S. F. (2018). The developing mental number line: Does its directionality relate to 5- to 7-year-old children's mathematical abilities? *Frontiers in Psychology*, *9*, Article 1142. <https://doi.org/10.3389/fpsyg.2018.01142>
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*(1), 125-135. <https://doi.org/10.1111/j.1467-7687.2010.00962.x>
- Barth, H., Slusser, E., Kanjlia, S., Garcia, J., Taggart, J., & Chase, E. (2016). How feedback improves children's numerical estimation. *Psychonomic Bulletin & Review*, *23*(4), 1198-1205. <https://doi.org/10.3758/s13423-015-0984-3>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear Mixed-Effects Models Using Eigen and S4 [Computer software] (R package version 1.1-7). Retrieved from <http://CRAN.R-project.org/package=lme4>
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*(2), 545-551. <https://doi.org/10.1037/a0017887>
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189-201. <https://doi.org/10.1037/0012-1649.41.6.189>
- Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*(4), 1016-1031. <https://doi.org/10.1111/j.1467-8624.2008.01173.x>
- Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology*, *44*, 1478-1490. <https://doi.org/10.1037/a0013110>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY, USA: Springer.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodal inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261-304. <https://doi.org/10.1177/0049124104268644>
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodal inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*, 23-35. <https://doi.org/10.1007/s00265-010-1029-6>

- Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on “Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures”. *Science*, 323, 38b. <https://doi.org/10.1126/science.1164773>
- Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 7867-7872. <https://doi.org/10.1073/pnas.1402785111>
- Cipora, K., & Nuerk, H.-C. (2013). Is the SNARC effect related to the level of mathematics? No systematic relationship observed despite more power, more repetitions, and more direct assessment of arithmetic skill. *Quarterly Journal of Experimental Psychology*, 66(10), 1974-1991. <https://doi.org/10.1080/17470218.2013.772215>
- Cipora, K., Patro, K., & Nuerk, H.-C. C. (2015). Are spatial-numerical associations a cornerstone for arithmetic learning? The lack of genuine correlations suggests no. *Mind, Brain and Education*, 9(4), 190-206. <https://doi.org/10.1111/mbe.12093>
- Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, 18(2), 331-338. <https://doi.org/10.3758/s13423-011-0059-z>
- Cohen, D. J., & Ray, A. (2020). Experimental bias in number-line tasks and how to avoid them: Comment on Kim and Opfer (2017) and the introduction of the Cohen Ray number-line task. *Developmental Psychology*, 56(4), 846-852. <https://doi.org/10.1037/dev0000761>
- Cohen, D. J., & Sarnecka, B. W. (2014). Children’s number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, 50(6), 1640-1652. <https://doi.org/10.1037/a0035901>
- Dackermann, T., Huber, S., Bahnmueller, J., Nuerk, H.-C., & Moeller, K. (2015). An integration of competing accounts on children’s number line estimation. *Frontiers in Psychology*, 6, Article 884. <https://doi.org/10.3389/fpsyg.2015.00884>
- Dackermann, T., Kroemer, L., Nuerk, H.-C., Moeller, K., & Huber, S. (2018). Influences of presentation format and task instruction on children’s number line estimation. *Cognitive Development*, 47, 53-62. <https://doi.org/10.1016/j.cogdev.2018.03.001>
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY, USA: Oxford University Press.
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, 320, 1217-1220. <https://doi.org/10.1126/science.1156540>
- De Vos, T. (1992). *TTR: Tempo Test Rekenen*. Lisse, The Netherlands: Swets & Zeitlinger.
- Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5-to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology*, 99(1), 1-17. <https://doi.org/10.1016/j.jecp.2007.08.006>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, 123, 53-72. <https://doi.org/10.1016/j.jecp.2014.01.013>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8, 307-314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Firebaugh, G. (2015). Statistics of ecological fallacy. In J. D. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., pp. 865–867). <https://doi.org/10.1016/B978-0-08-097086-8.44017-1>
- Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4, 59-65. [https://doi.org/10.1016/S1364-6613\(99\)01424-2](https://doi.org/10.1016/S1364-6613(99)01424-2)
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, 47, 1539-1552. <https://doi.org/10.1037/a0025510>
- Georges, C., Hoffmann, D., & Schiltz, C. (2017). Mathematical abilities in elementary school: Do they relate to number-space associations? *Journal of Experimental Child Psychology*, 161, 126-147. <https://doi.org/10.1016/j.jecp.2017.04.011>
- Gross, S. I., Gross, C. A., Kim, D., Lukowski, S. L., Thompson, L. A., & Petrill, S. A. (2018). A comparison of methods for assessing performance on the number line estimation task. *Journal of Numerical Cognition*, 4(3), 554-571. <https://doi.org/10.5964/jnc.v4i3.120>
- Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders’ knowledge of fractions. *Cognitive Development*, 35, 34-49. <https://doi.org/10.1016/j.cogdev.2015.02.001>
- Hoffmann, D., Hornung, C., Martin, R., & Schiltz, C. (2013). Developing number-space associations: SNARC effects using a color discrimination task in 5-year-olds. *Journal of Experimental Child Psychology*, 116(4), 775-791. <https://doi.org/10.1016/j.jecp.2013.07.013>

- Hoffmann, D., Mussolin, C., Martin, R., & Schiltz, C. (2014). The impact of mathematical proficiency on the number-space association. *PLoS One*, 9(1), Article e85048. <https://doi.org/10.1371/journal.pone.0085048>
- Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, 107(3), 500-524. <https://doi.org/10.1037/0033-295X.107.3.500>
- Hornung, C., Schiltz, C., Brunner, M., & Martin, R. (2014). Predicting first-grade mathematics achievement: The contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Frontiers in Psychology*, 5, Article 272. <https://doi.org/10.3389/fpsyg.2014.00272>
- Jung, S., Roesch, S., Klein, E., Dackermann, T., Heller, J., & Moeller, K. (2020). The strategy matters: Bounded and unbounded number line estimation in secondary school children. *Cognitive Development*, 53, Article 100839. <https://doi.org/10.1016/j.cogdev.2019.100839>
- Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. *Developmental Psychology*, 53(6), 1088-1097. <https://doi.org/10.1037/dev0000305>
- Link, T., Huber, S., Nuerk, H. C., & Moeller, K. (2014). Unbounding the mental number line—new evidence on children’s spatial representation of numbers. *Frontiers in Psychology*, 4, Article 1021. <https://doi.org/10.3389/fpsyg.2013.01021>
- Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *Quarterly Journal of Experimental Psychology*, 67(8), 1597-1613. <https://doi.org/10.1080/17470218.2014.892517>
- Luwel, K., Peeters, D., Dierckx, G., Sekeris, E., & Verschaffel, L. (2018). Benchmark-based strategy use in atypical number lines. *Canadian Journal of Experimental Psychology*, 72(4), 253-263. <https://doi.org/10.1037/cep0000153>
- Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. C. (2009). Children’s early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, 103(4), 503-515. <https://doi.org/10.1016/j.jecp.2009.02.006>
- Möhring, W., Frick, A., & Newcombe, N. S. (2018). Spatial scaling, proportional thinking, and numerical understanding in 5- to 7-year-old children. *Cognitive Development*, 45, 57-67. <https://doi.org/10.1016/j.cogdev.2017.12.001>
- Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified approach. *Cognition*, 149, 11-17. <https://doi.org/10.1016/j.cognition.2015.11.015>
- Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2016). Children’s use of number line estimation strategies. *European Journal of Psychology of Education*, 31, 117-134. <https://doi.org/10.1007/s10212-015-0251-z>
- Peeters, D., Verschaffel, L., & Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *British Journal of Psychology*, 108(4), 668-686. <https://doi.org/10.1111/bjop.12233>
- Petitto, A. L. (1990). Development of number line and measurement concepts. *Cognition and Instruction*, 7, 55-78. https://doi.org/10.1207/s1532690xci0701_3
- R Development Core Team. (2007). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reinert, R. M., Hartmann, M., Huber, S., & Moeller, K. (2019). Unbounded number line estimation as a measure of numerical estimation. *PLoS One*, 14(3), Article e0213102. <https://doi.org/10.1371/journal.pone.0213102>
- Reinert, R. M., Huber, S., Nuerk, H.-C., & Moeller, K. (2017). Sex differences in number line estimation: The role of numerical estimation. *British Journal of Psychology*, 108(2), 334-350. <https://doi.org/10.1111/bjop.12203>
- Rouder, J. N., & Geary, D. C. (2014). Children’s cognitive representation of the mathematical number line. *Developmental Science*, 17(4), 525-536. <https://doi.org/10.1111/desc.12166>
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, 114, 418-431. <https://doi.org/10.1016/j.jecp.2012.10.012>
- Sasanguie, D., Verschaffel, L., Reynvoet, B., & Luwel, K. (2016). The development of symbolic and non-symbolic number line estimations: Three developmental accounts contrasted within cross-sectional and longitudinal data. *Psychologica Belgica*, 56(4), 382-405. <https://doi.org/10.5334/pb.276>
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 20, Article 12372. <https://doi.org/10.1111/desc.12372>
- Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology*, 101, 359-372. <https://doi.org/10.1037/a0013840>

- Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., ... Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, 23, 409-422. <https://doi.org/10.1016/j.cogdev.2008.07.002>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, 89(5), 1467-1484. <https://doi.org/10.1111/cdev.13068>
- Sella, F., Berteletti, I., Lucangeli, D., & Zorzi, M. (2015). Varieties of quantity estimation in children. *Developmental Psychology*, 51(6), 758-770. <https://doi.org/10.1037/a0039183>
- Sella, F., Sader, E., Lollot, S., & Cohen Kadosh, R. (2016). Basic and advanced numerical performances relate to mathematical expertise but are fully mediated by visuospatial skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(9), 1458-1472. <https://doi.org/10.1037/xlm0000249>
- Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, 19(3), 341-361. <https://doi.org/10.1111/desc.12395>
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75, 428-444. <https://doi.org/10.1111/j.1467-8624.2004.00684.x>
- Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives*, 8, 144-150. <https://doi.org/10.1111/cdep.12077>
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237-243. <https://doi.org/10.1111/1467-9280.02438>
- Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, 62, 273-296. <https://doi.org/10.1016/j.cogpsych.2011.03.001>
- Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology*, 145, 22-33. <https://doi.org/10.1016/j.jecp.2015.12.004>
- Slusser, E., & Barth, H. (2017). Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks. *Journal of Experimental Child Psychology*, 162, 181-198. <https://doi.org/10.1016/j.jecp.2017.04.010>
- Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, 142(1), 193-208. <https://doi.org/10.1037/a0028560>
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 683-692. <https://doi.org/10.1037/0096-1523.16.4.683>
- Spence, I., & Krizel, P. (1994). Children's perception of proportion in graphs. *Child Development*, 65, 1193-1213. <https://doi.org/10.2307/1131314>
- Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, 18, 557-563. <https://doi.org/10.3758/s13423-011-0081-1>
- Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development*, 85, 1461-1476. <https://doi.org/10.1111/cdev.12218>
- White, S. L., & Szűcs, D. (2012). Representational change and strategy use in children's number line estimation during the first years of primary school. *Behavioral and Brain Functions*, 8(1), Article 1. <https://doi.org/10.1186/1744-9081-8-1>
- Yuan, L., Prather, R., Mix, K. S., & Smith, L. B. (2020). Number representations drive number-line estimates. *Child Development*, 91(4), e952-e967. <https://doi.org/10.1111/cdev.13333>



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.