

A potential dissociation between perception and production version for bounded but not unbounded number line estimation

Regina Miriam Reinert^a, Venera Gashaj^{a,b,*}, Matthias Hartmann^{c,d}, Korbinian Moeller^{b,e,f,g}

^a University of Tübingen, Tübingen, Germany

^b Centre for Mathematical Cognition, School of Science, Loughborough University, United Kingdom

^c Department of Psychology, University of Bern, Bern, Switzerland

^d Faculty of Psychology, UniDistance Suisse, Brig, Switzerland

^e Leibniz-Institut für Wissensmedien, Tübingen, Germany

^f LEAD Graduate School and Research Network, University of Tübingen, Germany

^g Individual Development and Adaptive Education Center for Children at Risk, Frankfurt am Main, Germany

ARTICLE INFO

Keywords:

Bounded/unbounded number line estimation tasks
Mental representation of number magnitude
Perception and production versions
Basic arithmetic skills

ABSTRACT

Background: What, exactly, do number line estimation (NLE) tasks measure? Different versions of the task were observed to have different effects on performance.

Method: We investigated associations between the production (indicating the location) and perception version (indicating the number) of the bounded and unbounded NLE task and their relationship to arithmetic.

Results: A stronger correlation was observed between the production and perception version of the unbounded than the bounded NLE task, indicating that both versions of the unbounded—but not the bounded—NLE task measure the same construct. Moreover, overall low but significant associations between NLE performance and arithmetic were only observed for the production version of the bounded NLE task.

Conclusion: These results substantiate that the production version of bounded NLE seems to rely on proportion judgment strategies, whereas both unbounded versions and the perception version of the bounded NLE task may rely more on magnitude estimation.

1. Introduction

Number line estimation (NLE), or locating a number on a given number line, is a common task used in research on numerical development to assess children's understanding of number magnitude (see [1] for a review and meta-analyses). There are different versions of the NLE task. The most applied version is the bounded NLE task by Siegler and Opfer [2]. In the vast majority of studies published on NLE (see e.g. [2–11]), a so-called production version of the bounded NLE task was used, where the position of a given number has to be estimated on a number line with specified start and endpoint numbers (henceforth *production* version, cf. [1]). As argued by Siegler and Opfer [2], conclusions about the underlying representation of number magnitude can be inferred directly from participants' estimation pattern, which is best explained by linear or logarithmic response functions (see e.g. [12, 4,13,9], but see Barth and Palladino [14] for differing results). Furthermore, the repeated use of this task, therefore, should allow to

track the development of children's representation of number magnitude over time (e.g., [15]). With experience and age, the estimation pattern was found to change from an initially less accurate logarithmic estimation pattern to a more precise and linear one (e.g. [5,16,17]).

In some studies, this typical production version of the bounded NLE task is complemented by a *perception* version in which participants have to indicate which number best describes an already marked position on a number line ([2], henceforth perception version of the task). However, only few studies have used this version of the task so far [2,18–22]. In the seminal study by Siegler and Opfer [2], the authors noted linear estimation patterns in the production version of the bounded NLE task in a sample of second graders on a 0 to 100 scale, but logarithmic ones on a 0 to 1,000 scale. Importantly, considerably different estimation patterns were observed in the perception version of the bounded NLE task where the same second graders showed estimation patterns best described by an exponential function for both scales. However, older students as well as adults also showed linear performance patterns in the perception

* Corresponding author at: Loughborough University, Leicestershire LE11 3TU, United Kingdom.

E-mail address: V.Gashaj@lboro.ac.uk (V. Gashaj).

<https://doi.org/10.1016/j.tine.2023.100202>

Received 16 November 2022; Received in revised form 5 April 2023; Accepted 12 April 2023

Available online 16 April 2023

2211-9493/© 2023 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

version of the bounded NLE task. Hence, these results provided first evidence that these two task versions might not measure the same underlying construct.

There is an ongoing debate with respect to the validity of the traditional bounded NLE task as it might not capture pure numerical estimation (e.g. [14,23–30]). In particular, Barth and colleagues [14,31,32] argued that changes in estimation patterns may be better described by an increasing reliance on proportion judgements, rather than by a logarithmic-to-linear representational shift. Completing the NLE task successfully seems to depend on the application of appropriate strategies (see e.g. [14,24,32–37]). There is accumulating evidence that children estimate target numbers based on reference points (e.g., the start and endpoint as well as the middle of the number line). Hence, logarithmic-to-linear changes in response patterns may originate from constraints that are task specific (see [25]).

Cohen and Blanc-Goldhammer [24] introduced the unbounded number line estimation task to address this issue. Only a standard line segment with a start point and a scaling unit with two vertical lines depicting a unit (usually 1) are provided in this task version. In the production version, individuals are asked to estimate the magnitude of a presented target number based on this unit length. For the bounded NLE task, an M-shaped pattern of estimation errors smaller and less variant at and around reference points is typically observed. In contrast, the distribution of errors is increasing linearly for the unbounded NLE task and thus seems to overcome the limitations of the traditional bounded task version. The unbounded version is therefore considered to be a purer and more valid measure of number magnitude estimation (see [21,24,25,35–38]; for a review see [39]).

As this task variant is fairly new, only two studies have employed a perception version of the unbounded NLE task so far [19,21]. Cohen et al. [19] found that the perception version of the unbounded NLE task produced the same estimation biases as the production versions of both the bounded and the unbounded NLE task. Only the perception version of the bounded NLE task was exceptional showing systematic biases different from the three other task versions. In particular, participants overestimated the quantity presented in the perception version of the bounded NLE task more than its production counterpart. Conversely, Reinert et al. [21] observed systematic patterns of underestimation in the perception version of the unbounded NLE task as well as in the perception version of a dot estimation task. This result corroborates the notion that the unbounded NLE task seems conceptually similar to the non-symbolic numerosity estimation task and may thus represent a purer measure of number magnitude estimation.

1.1. Associations of bounded and unbounded NLE with arithmetic skills

There is accumulating evidence to question whether bounded and unbounded NLE tasks measure the same underlying construct. Previous studies found differing correlations between different versions of the NLE task and broader mathematical outcomes (for a review, see [40]). More accurate estimations in the production version of the bounded NLE task were shown to be predictive of better mathematical achievement [18,41–43]. For example, Sasanguie et al. [44] observed that more linear estimation patterns in the production version of the bounded NLE task were associated with higher mathematic achievement (see also [12,45,46,7]). Furthermore, the study of Torbeyns et al. [43] indicated strong correlations between mathematical achievement scores of sixth and eighth graders and a bounded NLE task on fractions. Importantly, these associations remained significant in several studies, even when controlling for plausible confounding variables such as intelligence, executive function [47,4], parental income and education, gender [47], or reading achievement [48]. Schneider et al. [1] substantiated these associations in a meta-analysis observing correlations between various even more complex and advanced mathematical competence measures and bounded NLE performance. Hence, there is considerable evidence that the production version of the bounded NLE task is a meaningful

predictor of broader mathematical skills (see also [1,40]).

In contrast, there are only a few research studies on correlations between unbounded NLE and math performance so far. From the 17 existing studies on unbounded NLE, only four evaluated the association of unbounded NLE performance and other numerical and mathematical skills [49,38,50,27]. For instance, Link et al. [27] observed no significant associations between unbounded NLE and numerical or mathematical skills in fourth-graders, but significant correlations between bounded NLE and several numerical and mathematical tasks (including addition and subtraction performance). Similarly, Georges and Schiltz [49] also found no significant association between the production version of unbounded NLE and addition as well as subtraction skills in second and fourth graders (but did so for bounded NLE, see [50], for differing results). Thus, previous studies mostly did not substantiate the association of unbounded NLE with other numerical/mathematical skills.

1.2. Strategy use in bounded and unbounded NLE

In sum, the findings described above suggest that bounded and unbounded NLE might not measure the same underlying construct as at least the traditional bounded NLE task seems to be confounded by specific strategies used to solve this task (e.g. [26]). The significant association with other numerical and mathematical skills observed in previous studies suggests that estimation performance in bounded NLE may be driven by proportional judgement strategies. More specifically, participants may use reference points such as the midpoint of the number line to complete this task version, rather than mere estimation (e.g. [51]). In contrast, the unbounded NLE task seems to involve more pure estimation strategies and may therefore be more independent of specific estimation strategies such as proportion judgement. Individuals were instead found to develop their own strategies to solve this task version (see [24,35,52]).

1.3. The current study

The purpose of the present study was to investigate (i) the association of perception and production version of bounded and unbounded NLE tasks and (ii) their potentially selective association with primary school children's performance on a variety of different basic numerical and mathematical skills. To the best of our knowledge, so far, there is no study evaluating similarities and differences between perception and production versions of bounded and unbounded NLE in terms of their respective correlations but also their associations with other numerical and mathematical skills. Doing so will provide further evidence on whether NLE tasks measure the same underlying construct depending on task version (bounded vs. unbounded) and presentation format (perception vs. production version).

One may conjecture that perception and production versions of bounded and unbounded NLE reflect the same task demands (i.e., processing spatial-numerical correspondence in a setting allowing proportion judgement (bounded) or not (unbounded NLE)). Accordingly, we expected significant correlations between the perception and the production version of the bounded and unbounded NLE tasks, respectively. Moreover, in line with previous studies, we expected to replicate the significant association of the (mostly employed) production version of the bounded NLE task with other basic numerical and arithmetic skills as completion (see [12,45,27,37]). In contrast, no such correlation is expected for production version of unbounded NLE according to previous studies (e.g. [49,27]). Based on the expected correlation between the perception and the production versions of bounded and unbounded NLE, a similar pattern of associations was expected for the perception version of the bounded NLE task.

2. Methods

2.1. Participants

A total of 142 fourth graders (68 girls) with a mean age of 10.4 years ($SD = 5.6$ months) participated in this study. They were recruited from ten different public elementary schools in Switzerland and assessed on a battery of tasks evaluating basic arithmetic skills (see below for a more detailed description). All children were German speaking and participated voluntarily. Prior to testing, written informed consent was obtained from children's parents and the headmaster, while oral assent was obtained from children and teachers. The study was approved by the local ethics committee of the University of Bern (Nr. 2017-10-00003).

2.2. Stimuli and procedure

The present study was carried out in the context of a larger research project that also investigated fine and gross motor skills, working memory, executive functions such as inhibition and switching, self-concept, and physical activity. However, for the current research question only data from bounded and unbounded NLE as well as children's basic arithmetic performance as assessed by a standardized math achievement test battery (Heidelberger Rechentest, HRT; [53]) were considered.

All children completed the different tasks in two sessions during regular school hours. About half of the pupils started with a classroom session, including administration of the HRT which was conducted in paper-based form as per manual. A booklet including all test materials was distributed to keep the order of tasks constant for all participants. This also comprised assessment of background variables (e.g., demographics, SES, etc.), questionnaires evaluating physical activity and self-concept. In case children finished a task or questionnaire earlier, they were told to wait silently and not to begin with the next task / questionnaire until instructed. The entire group session lasted about 45 min. Completion of the six HRT subtests (i.e., Addition, Subtraction, Multiplication, Division, Number Sequences, and Number Comparison) took about 15 min.

In the other session, children performed the computerized NLE tasks as well as tasks assessing executive functioning (measuring working memory, inhibition, and switching) individually in a separate room. Stimuli of both the unbounded and bounded NLE tasks were presented as pictures on a laptop with a 15.4' WXGA screen at a resolution of $1,280 \times 800$ pixels. Both NLE tasks included a perception and a production version of the respective task and children were asked to respond as fast and as accurate as possible. The order of all computerized tasks varied pseudo-randomly with the constraint that students never completed an unbounded NLE task directly following a bounded NLE task. This was done to avoid students transferring the number range from the bounded to the unbounded NLE task. Furthermore, children did not receive any information about the range of numbers used in the unbounded NLE task. For both task versions, no feedback was provided as to the correctness of children's responses. For each participant, target numbers were presented in randomized order. Performing the NLE tasks in the individual session took about half an hour. The other half of children completed the experiment with sessions reversed.

2.3. Number line estimation

In the *perception version* (position-to-number) of both the bounded as well as the unbounded NLE tasks, a target position was already marked with a blue vertical line on the number line. Children were asked to insert the Arabic number that is reflected by the respective spatial location using the number keys of the laptop keyboard. A box in which their response was shown was displayed above the start point and children had to press the "Enter" button to login their final response.

In the *production version* (number-to-position) of the bounded and

unbounded NLE tasks, participants were given a target number and requested to indicate its spatial position on an otherwise empty number line using the mouse to click at the estimated position. The blue vertical line with which they had to indicate the estimated location of the target always appeared in the center of the screen (for the unbounded version) or at the starting point (for the bounded version) on the number line and had a vertical length of about 1.5 cm. All number lines and target numbers were displayed in black color against a white background (see Fig. 1) on a Lenovo 3000 N200 laptop with a screen size of 15.4', aspect ratio of 16:10, and driven at a resolution rate of $1,280 \times 800$ pixels.

The *bounded number line estimation task* covered the number range from 0 to 1,000. A total of sixteen items was displayed both in the perception as well as in the production version of the task [13, 24, 67, 125, 234, 285, 363, 426, 517, 586, 671, 736, 834, 916, 981, 997]. Target numbers were equally distributed across the whole range with a slight oversampling at around the start, mid- and endpoint of the scale as potential landmarks. The number line was displayed at the same position on screen for all items at a physical length of 19.5 cm.

Two different sets of fifteen target numbers, ranging from 2 to 49, were created for the *unbounded number line estimation task*: one for the perception version [3, 4, 9, 12, 15, 19, 23, 24, 28, 31, 35, 36, 43, 46, 48] and another for the production version [2, 5, 8, 13, 14, 18, 21, 25, 26, 32, 37, 38, 42, 45, 49]. Target numbers were chosen to be on average equally distributed across the entire number range covered. As in previous studies, a smaller number range was used for the unbounded version, as there is no evidence for an influence of the number range on participants' estimation performance (see [24,50,35,52]). Number lines had a numerical length of 50 units (with 48 being the largest target number in the perception and 49 in the production version) at a physical length of 19.4 cm. The size of unit 1 was indicated at the start point reflecting a segment with a length of 0.3 cm. The position of the number lines varied randomly on the screen to prevent children from using external reference points.

For both bounded as well as unbounded NLE the *percentage absolute error* [henceforth PAE = $|\text{estimated number} - \text{target}|/\text{scale}$; cf. [9]] was used as dependent variable reflecting children's NLE performance. Thereby, we standardized children's estimation errors on the number range of both number line tasks to ensure comparability.

Standard Math Achievement Test (Heidelberger Rechentest; HRT 1–4). Five subtests of the HRT were used to assess children's basic arithmetic abilities: i) addition (e.g., "16 + 27 = _"), ii) subtraction (e.g., "50 - 14 = _"), iii) multiplication (e.g., "8 × 17 = _"), iv) division (e.g., "28 : 4 = _"), and v) number comparison (e.g., "2 + 9 • 20" Is 2 + 9 bigger (>) or smaller (<) than 20 ?). Each of these subtests consists of 40 items of increasing difficulty, beginning with simple items. Additionally, the subtest (vi) number sequences (e.g., "3 3 4 5 6 _ _") was administered, in which children have to add the next three numbers to given number sequences by identifying and extending the logical relationship between numbers. The HRT is designed as a speed test to solve as many trials as possible within the two minutes time limit for the arithmetic and three minutes for the number sequences subtests. In this study, sum scores of correctly solved items served as the dependent variable. The maximum score for each subtest of the arithmetic part is 40 points, whereas 20 points can be achieved in the number sequences subtest.

3. Results

Regarding NLE tasks, estimates that differed more than ± 3 standard deviations from individuals' overall mean estimates for the respective task version were excluded from further analyses in a first step. This resulted in a loss of 0.01% of the data. Moreover, mean estimates of a single subtask were missing for two children. Overall, 20 children did not participate in the group sessions in which the HRT was administered, so our analyses were based on data of at least 121 children. However, more data sets were available for the analyses on NLE tasks that were collected in the individual sessions (see Table 1).

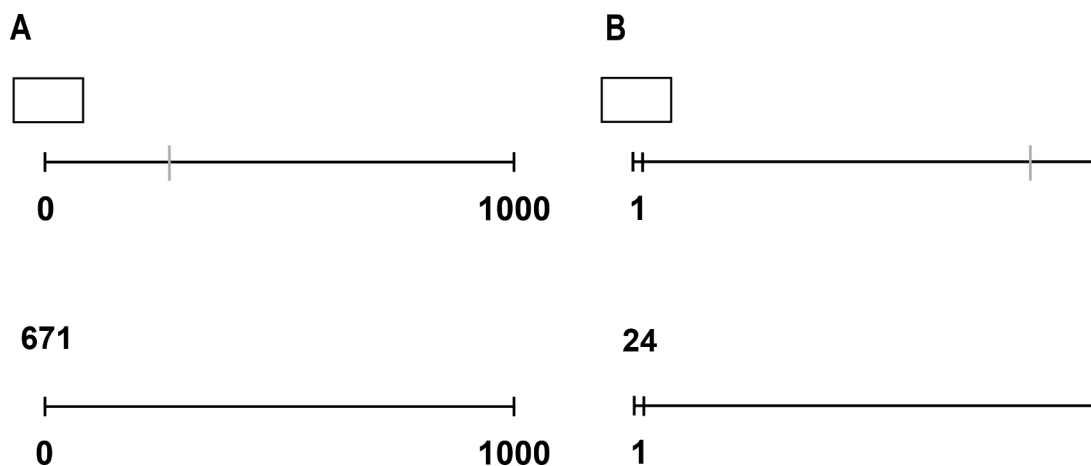


Fig. 1. Bounded and unbounded NLE tasks.

Note. Example of an (A) bounded and (B) unbounded NLE task. The perception version is depicted in the upper row, the production version in the lower row.

Table 1
Descriptive information.

Variable	N	M	SD	Range (Min - Max)
Sex (female/male)	68/ 74	-	-	-
Age (years)	142	10.36	0.47	9.30 – 11.87
Bounded number line – perception (PAE)	137	3.28	3.02	0.03 – 15.49
Bounded number line – production (PAE)	140	3.54	3.08	0.02 – 14.94
Unbounded number line – perception (PAE)	142	25.75	14.53	0.53 – 70.53
Unbounded number line – production (PAE)	141	16.52	8.79	0.06 – 41.11
HRT Addition (score)	122	27.17	4.31	18 – 37
HRT Subtraction (score)	122	25.11	5.65	14 – 37
HRT Multiplication (score)	122	21.16	6.21	8 – 32
HRT Division (score)	122	18.89	8.39	1 – 31
HRT Number comparison (score)	121	26.88	6.36	6 – 40
HRT Number sequences (score)	122	11.93	2.06	6 – 16

3.1. Descriptive statistics

All descriptive information is depicted in Table 1.

3.2. Correlation analysis

An overview of all raw correlations between the variables of interest is given in Table 2. As expected, all HRT subtests were significantly inter-related.

Table 2
Correlation analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Bounded number line – perception (PAE)	-	.273**	-0.009	-0.071	-0.082	-0.163⁺	-0.036	-0.095	-0.087	-0.177⁺
(2) Bounded number line – production (PAE)		-	.049	-0.007	-0.186*	-0.203*	-0.153⁺	-0.154⁺	-0.165⁺	-0.180*
(3) Unbounded number line – perception (PAE)			-	.569***	-0.027	-0.069	.009	.011	.071	-0.065
(4) Unbounded number line – production (PAE)				-	-0.134	-0.085	-0.034	-0.045	-0.032	-0.064
(5) HRT Addition (score)					-	.721***	.574***	.559***	.594***	.546***
(6) HRT Subtraction (score)						-	.664***	.620***	.687***	.647***
(7) HRT Multiplication (score)							-	.758***	.500***	.554***
(8) HRT Division (score)								-	.576***	.564***
(9) HRT Number comparison (score)									-	.561***
(10) HRT Number sequences (score)										-

Note. ⁺*p* < .10 **p* < .05. ***p* < .01. ****p* < .001. Bold script indicates correlations between different versions of the number line estimation task and basic numerical and arithmetical skills.

3.2.1. Correlations between basic arithmetic skills and number line estimation

Overall, only few significant correlations were found between NLE and arithmetic performance. Importantly, and in line with previous evidence, no significant correlations with arithmetic tasks were observed for both versions of the unbounded NLE task. Interestingly, this was also the case for the perception version of the bounded NLE task.

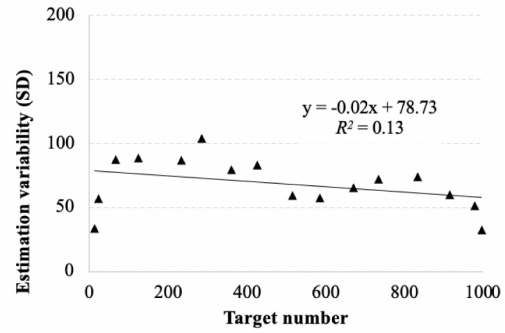
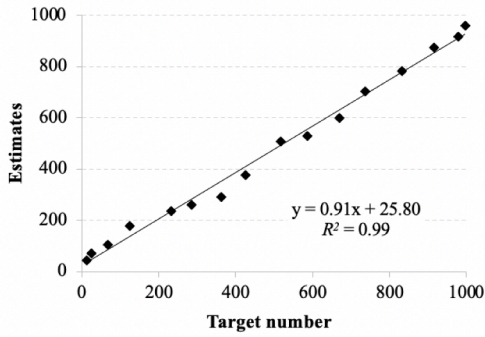
Significant associations between NLE and arithmetic tasks were only observed for the production version of the bounded NLE task. In line with previous studies (see [12,45,5,27]), children’s estimation accuracy (in terms of PAE) correlated significantly with addition and subtraction performance as well as performance on the number sequences subscale. For all associations, correlation coefficients indicated that children with more accurate estimates in the production version of the bounded NLE task attained a higher score on the respective HRT subtest.

3.2.2. Bounded and unbounded number line estimation

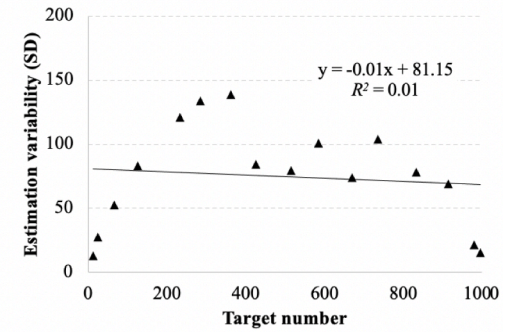
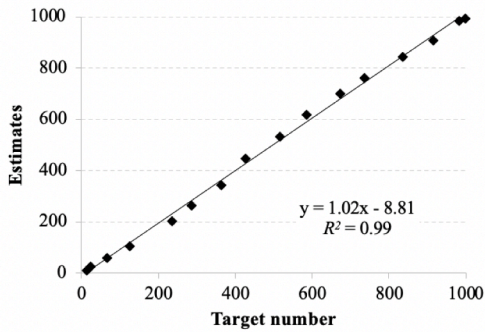
Mean estimates and estimation errors were plotted as a function of target number (see Fig. 2) to evaluate potential differences in fourth graders’ bounded and unbounded NLE performance. For these charts, all individual estimates that differed more than ± 3 standard deviations from groups’ mean estimate of the respective target number were excluded from further analyses. This resulted in a loss of 0.03% of trials in both the production and the perception version of the bounded NLE task. In the unbounded NLE tasks, this procedure led to a loss of 0.008% of trials in the production and 0.009% in the perception version.

The distribution of errors, however, showed that participants evidently used reference points in both the production as well as in the perception version of the bounded NLE task: The characteristic M-shaped pattern in the right column of Panels A and B indicates more minor

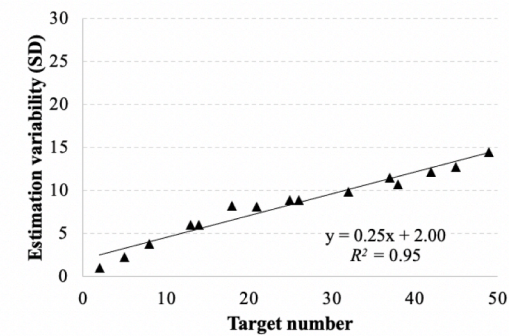
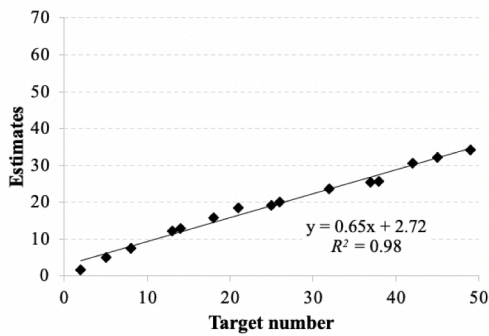
A Bounded number line - Production version



B Bounded number line - Perception version



C Unbounded number line - Production version



D Unbounded number line - Perception version

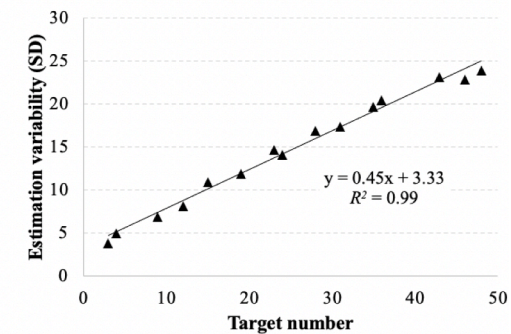
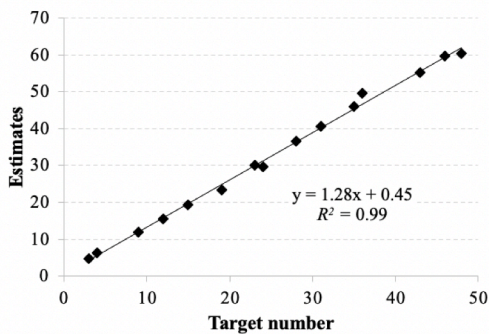


Fig. 2. Estimation patterns for bounded and unbounded NLE

Note. Estimation patterns (mean estimates across all participants (left charts) and variability of estimation errors (SD of PAE, right charts) for bounded (Panels A and B) and unbounded NLE (Panels C and D)).

estimation errors at and around the start, mid- and endpoint (i.e., 0, 500, and 1,000). In line with Ashcraft and Moore's [18] procedure, we conducted a contour analysis contrasting children's estimation errors at and around reference points with those farthest away from common

reference points in terms of absolute estimation error (see also [37]). For these analyses, we considered all target numbers ± 25 around the three reference points. This means the two target numbers nearest the start point (i.e., 13 and 24), the midpoint (i.e., 517), and the two observations

closest to the endpoint (i.e., 981 and 997) as well as those target numbers farthest away from reference points (i.e., 234 and 736).

A 2×2 repeated-measures ANOVA with the factors *task version* (production vs. perception version of the bounded NLE task) and *distance from reference point* (close to vs. far away from reference points) was conducted to evaluate whether the characteristic M-shaped error pattern indeed reflected significantly smaller estimation errors at and around reference points in both versions of the bounded NLE task.

The ANOVA revealed a main effect of distance from reference point indicating that the mean absolute estimation error was significantly smaller around the reference points (i.e., 0, 500, and 1,000) compared to those target numbers farthest away (i.e., 250 and 750), $M_{close} = 32.8$ versus $M_{far\ away} = 69.6$; $F(1, 116) = 186.35$, $p < .001$, $\eta_p^2 = 0.616$. In contrast, no significant main effect of task version was observed, $F(1, 116) = 0.38$, $\eta_p^2 = 0.003$, suggesting that children may have used proportion-judgement strategies to solve both task versions resulting in an M-shaped error pattern, $M_{production} = 50.3$ versus $M_{perception} = 52.1$. At last, the ANOVA revealed a significant interaction of *task version* and *distance from reference points*, $F(1, 116) = 38.06$, $p < .001$, $\eta_p^2 = 0.247$, which indicated that advantage for the estimates at and around reference points was more pronounced in the perception version compared to the production version of the bounded NLE task (18.9 versus 54.7, respectively, see Fig. 3).

In contrast, estimations errors in both unbounded NLE task versions increased linearly with target size ($r \geq 0.95$). Additionally, fourth graders made more accurate estimates in the production than the perception version of the unbounded NLE task.

3.2.3. Correlation between the perception and production version

As depicted in Table 2, significant correlations of children's estimation accuracy – as indicated by PAE – were found between the perception and production version of bounded $r(135) = 0.273$, $p = .001$ as well as unbounded NLE $r(141) = 0.569$, $p < .001$. To evaluate whether the association of perception and production versions differed significantly between bounded and unbounded NLE tasks, Steiger's Z-Test [54] was used. We observed that the correlation between the perception and production version of the bounded NLE task was significantly lower than the correlation between these two task versions of the unbounded NLE task [$Z(131) = -2.97$, $p = .003$, tested two-sided].

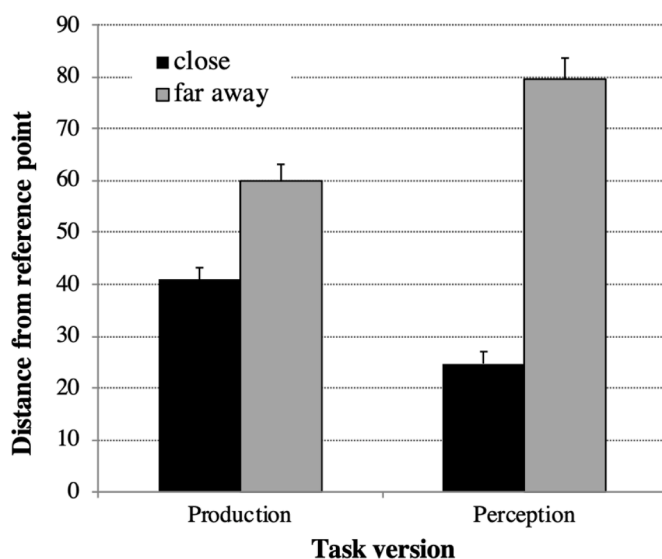


Fig. 3. Estimation errors at and around reference points in the different NLE tasks
Note. Marginal means of absolute estimation errors for production and perception version of the bounded NLE task separated for close areas at reference points and those more far away. Error bars reflect 1 SEM.

Hence, Steiger's Z-Test indicated a significantly stronger association between the perception and production version for the unbounded NLE task than the bounded NLE task.

4. Discussion

The present study aimed at better understanding what bounded and unbounded NLE tasks measure. Therefore, we systematically evaluated differential associations between the perception and production version of bounded and unbounded NLE and their association with other arithmetic tasks in 9- to 11-year-old primary school students. To our knowledge, this is the first study evaluating potential associations and dissociations across these task versions in this age group. In line with previous studies [12,45,5,27], we observed significant associations between basic arithmetic performance and NLE for the production version of the bounded NLE task, but not for the perception version of the bounded nor any version of the unbounded NLE task versions. Moreover, a significantly stronger correlation was observed between the perception and production version of the unbounded compared to the traditional bounded NLE task versions. In the following, we will discuss these findings in more detail.

4.1. Differential associations with basic arithmetic skills

In line with previous evidence (see [12,45,5,27]), we also found estimation performance in the production version of the bounded NLE task to be significantly related to most of the arithmetic tasks assessed. In particular, significant correlations were observed between NLE accuracy and addition, subtraction, as well as performance in the number sequences task.

We did not observe any significant association between the perception version of the bounded NLE or the two versions of the unbounded NLE and mathematical skills. The observed correlations between mathematical skills and the two versions of the bounded and unbounded task differed considerably. We observed that only the production version, but not the perception version, of the bounded NLE task correlated significantly with mathematical skills. This indicates that participants seem to apply different solution strategies to complete these versions of the bounded NLE task. One might note that children are more familiar with the production version of the bounded NLE task than with its perception version as well as the unbounded NLE task in general. However, we standardized estimation errors on the number range covered by the respective task versions and did focus on correlations with other arithmetical tasks and not on differences in task performance. As such, we are confident that potential differences in task difficulty should not bias the results.

Beyond that, these differential association patterns for all four versions of the NLE task with arithmetic skills suggest that the production version of the bounded NLE task may be solved by applying classroom (learnt) strategies that involve specific arithmetic operations, such as addition and subtraction. For instance, when requested to locate the target '691' on a 0 to 1,000 number line, children may first consider '500' as the midpoint of the number line halfway between the lower '0' and upper '1,000' bound. They may start quartering it ('250' and '750'). Next, they may evaluate whether the target '691' is smaller or larger than a quarter/midpoint and compute the distance by adding ('500' + 191') or subtracting ('750' - '59') some units (see also [55] for a more detailed discussion, see also [26]).

Besides these associations with simple arithmetic operations, the significant correlation between the production version of the bounded NLE task and the HRT subtest *number sequences* indicates an association of NLE and more complex numerical processes. In this subtest, children were asked to deduce the rule which governs the sequence and derive the next three numbers of the series following its logical structure. Participants needed to combine several arithmetic operations in a row and perform addition and subtraction operations according to the

logical relationship between the numbers. For instance, they had to complete the number sequence ‘3 3 4 5 6 _ _ _’ (see [53]) by identifying that the rule is report odd numbers twice and even numbers once. Accordingly, the three numbers to be added are ‘7 7 8’ by combining these numerical-logical addition rules. Children who quickly understood the logic behind these sequences also showed more accurate estimates in the production version of the bounded NLE task. The example described above placing ‘691’ on the 0 to 1,000 line illustrates the combination of consecutive arithmetic and numerical operations showing how basic arithmetic as well as other numerical processes (e.g., magnitude comparison) contribute to estimation performance in the production version of the bounded NLE task.

Unlike Link et al. [27], we did not observe a significant correlation with the subtest *number comparison*. However, there is a tendency for this association [$r(120) = -0.17, p = .07$]. The lack of associations with multiplication and division seems consistent with the literature (e.g. [27]). Moreover, according to the model fittings in Cohen and Sarnecka’s [26] study, multiplication and/or addition was argued to be used by children for unbounded NLE, a fact which we also did not observe potentially indicating that children used addition predominantly. Cohen and Sarnecka [26] also assumed division and/or subtractions to solve the production version of the bounded NLE task. However, we did not observe significant associations of division with estimation performance for this task version in our study. The fact that we only observed significant associations of bounded NLE and subtraction might indicate that children used subtraction strategies predominantly.

In sum, these results seem to add to accumulating evidence suggesting that the production version of the bounded NLE task may not reflect an unbiased measure of numerical estimation but seems to measure arithmetic skills (i.e., addition and subtraction) besides the intended estimation processes. Thereby, our results substantiate the interpretation by Link et al. [27] who claim that the associations between the production version of the bounded NLE task and other basic numerical/arithmetic competencies seem to be driven by numerical/arithmetical processes required to successfully apply proportion judgement strategies. Estimation patterns obtained in this task version may therefore not reflect the underlying number magnitude representation directly as argued by Siegler and Opfer [2], but may be confounded with other numerical/arithmetical processes and might thus be biased. This seemed to be less pronounced for the other three task versions.

However, when interpreting these data, there is a crucial point to consider. While above considerations describe the observed correlation pattern, it must be noted that in particular correlations of bounded NLE with arithmetic tasks were considerably smaller than observed in previous studies like Link et al. [27]. As such, some caution is advised when arguing with these results. Also, some correlations between the perception version of the bounded NLE task and arithmetic approached significance (e.g., for number sequences). Considering this, these data only provide first evidence suggesting potential differential associations of arithmetic tasks with bounded and unbounded NLE and between the perception and production version – at least for bounded NLE. In essence, additional studies are needed to further evaluate and substantiate these results before drawing firm conclusions. Nevertheless, our results were more informative concerning different estimation patterns for the different NLE task versions.

4.2. Estimation patterns in the different number line estimation task versions

Regarding estimation patterns, the distribution of errors was informative regarding the solution strategies applied to solve the different task versions. It revealed that fourth-graders evidently used systematic reference points in both versions of the bounded NLE task. We observed the typical M-shaped error pattern in the production as well as in the

perception version of the bounded NLE task reflecting smaller estimation errors at and around typical reference points (i.e., the origin, mid- and endpoint, 0, 500, and 1,000, respectively). This pattern of results substantiated previous findings on the production version of the bounded but provides new evidence for the unbounded NLE task (e.g. [14, 36]). This indicates that in both bounded NLE task versions participants applied proportion judgement strategies to complete the tasks. Interestingly, we found larger estimation errors for the production compared to the perception version. This is in line with previous research indicating that the perception version of the bounded NLE task was substantially more demanding [20] which means that participants’ estimates were less accurate translating a given spatial position on a number line into a numerical value than inferring the spatial location of a given number on a number line.

In line with previous studies, estimation errors in both unbounded NLE task versions linearly increased with the size of the target number. Moreover, the estimates of children were more accurate in the production than in the perception version. The more linear distribution of estimation errors (and thus the lack of the characteristic M-shaped distribution) indicates that unbounded NLE primarily relies on numerical estimation rather than strategies associated with proportion judgement. However, it is important to note that we did not find any significant association between unbounded NLE and children’s arithmetic performance. In line with the argument by Reinert et al. [37], this suggests that strategies applied in unbounded NLE seem less dependent on classroom-learned methods and procedures. Instead, unbounded NLE seems to draw on spatial-numerical estimation.

Hence, systematic evaluation of production and perception versions of the bounded and unbounded number line estimation task provided for the first time converging evidence on the limitations of the bounded NLE task due to biases in solution strategies employed (see also [30,56–58]).

4.3. Consistency of both the perception and production versions of the number line estimation tasks

The results of the present study are also meaningful considering the differential correlations between the perception and production version of both the bounded and unbounded NLE tasks. In particular, we observed a significantly stronger correlation between the perception and production version of the unbounded compared to the traditional bounded NLE task. Compared to the two complementary versions of the unbounded NLE task, the significantly lower association of these two inverse task versions suggests that it makes a difference which version of the two bounded ones is used to interpret the respective estimation pattern. The percentage of explained variance of the association of the production and perception version of the bounded NLE task was 7.5% indicating that these may not measure the same.

This further corroborates the assumption that different solution strategies are recruited in these two task versions. The lack of associations between the perception task version and any arithmetic operation compared to the correlations between the production version of the bounded NLE task and several arithmetic operations confirms this claim (but see the limitations mentioned above on the size of the correlations overall). Nevertheless, this indicates that participants may apply different solution strategies in the perception version of the bounded NLE task. More specifically, our data suggest that this task version is solved more by estimating or counting-based strategies than by typical proportion judgement.

On the other hand, the percentage of explained variance shared between both unbounded NLE task versions is significantly higher (32%). Considering results of a study we conducted recently [21], indicated a similar result pattern with 18% of variance shared between the two bounded NLE task versions and a somewhat higher percentage of explained variance (30%) confirming the stronger association between the two complementary unbounded NLE task versions. This means that both unbounded NLE task versions seem to be solved by more similar

solution strategies that may be related to classroom-learned strategies such as arithmetic operations less but more so by mere estimation. Therefore, the two inverse versions of the unbounded NLE task capture number magnitude estimation more comparably, whereas this is less clear for the two bounded task versions.

5. Conclusions & perspectives

In summary, we observed first evidence suggesting significant associations with basic arithmetic skills seem more prominent for the production compared to the perception version of the bounded NLE task, and both unbounded NLE task versions. Furthermore, the perception and production version of the unbounded NLE task correlated significantly stronger than the two versions of the bounded NLE task. These findings support the claim that the production version of the bounded NLE task seems to be specific in terms of what it measures – which may not necessarily be numerical estimation – but solution strategies such as proportion judgement, which also involve arithmetic procedures. However, the overall low correlations between bounded NLE and basic arithmetic tasks observed in this study must be considered when interpreting these differential results. For the unbounded NLE task, in contrast, our results suggest that perception and production versions were solved more similarly. Future studies are needed to evaluate further and substantiate this initial evidence on influences of the classical production vs. perception version of the (bounded) NLE task. This seems relevant as number lines are a common tool for familiarizing children with different ranges of number magnitudes (e.g., 0 to 100 in second grade) and facilitating arithmetic operations within the respective number range. The present evidence suggests using the production version of the bounded NLE task in such educational settings as it is the one most strongly associated with children's more general arithmetic and mathematical performance. Additionally, further evidence indicates interventions building on bounded NLE to successfully increase children's arithmetic/mathematical performance (e.g. [59,60]). Notably, there was no such association for the perception version of the bounded as well as both unbounded NLE task versions. Yet, as the latter seems to reflect a purer measure of number magnitude representation, one might assume that the unbounded NLE task may be useful not only to assess but also to foster the representation of number magnitude. To evaluate this claim, future intervention studies using the unbounded NLE task would be desirable to investigate potential effects on children's number magnitude representation and as a consequence on the development of their mathematical skills more broadly. This would substantiate that both task versions assess different aspects of numerical skills.

Author Contributions

RMR Conceptualization, Methodology, Data Analysis, Writing-Original draft. **VG** Project Administration, Conceptualization, Data Acquisition, Review and Editing. **MH** Reviewing and Editing. **KM** Conceptualization, Resources (Analysis Tools, Tasks), Writing-Reviewing and Editing. All authors interpreted and discussed the results, reviewed, edited, and approved the last version of the manuscript.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

We are grateful to all those with whom we have enjoyed working during this project. We would like to thank all participating schools, teachers, 4th graders, and their parents. We also appreciate the help of our coworkers at the University of Bern with data collection and

technical support. Furthermore, we would like to thank Nicole Oberer and Dragan Trninić for their topical discussions, logistic support, and spell check.

References

- [1] M. Schneider, S. Merz, J. Stricker, B. De Smedt, J. Torbeyns, L. Verschaffel, K. Luwel, Associations of number line estimation with mathematical competence: a meta-analysis, *Child Dev.* 89 (5) (2018) 1467–1484, <https://doi.org/10.1111/cdev.13068>.
- [2] R.S. Siegler, J.E. Opfer, The development of numerical estimation: evidence for multiple representations of numerical quantity, *Psychol. Sci.* 14 (3) (2003) 237–243, <https://doi.org/10.1111/1467-9280.02438>.
- [3] I. Berteletti, D. Lucangeli, M. Piazza, S. Dehaene, M. Zorzi, Numerical estimation in preschoolers, *Dev. Psychol.* 46 (2) (2010) 545–551, <https://doi.org/10.1037/a0017887>.
- [4] D.C. Geary, M.K. Hoard, L. Nugent, J. Byrd-Craven, Development of number line representations in children with mathematical learning disability, *Dev. Neuropsychol.* 33 (3) (2008) 277–299, <https://doi.org/10.1080/87565640801982361>.
- [5] E.V. Laski, R.S. Siegler, Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison, *Child Dev.* 78 (6) (2007) 1723–1743, <https://doi.org/10.1111/j.1467-8624.2007.01087.x>.
- [6] K. Moeller, S. Pixner, L. Kaufmann, H.C. Nuerk, Children's early mental number line: logarithmic or rather decomposed linear? *J. Exp. Child Psychol.* 103 (4) (2009) 503–515, <https://doi.org/10.1016/j.jecp.2009.02.006>.
- [7] G.B. Ramani, R.S. Siegler, Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games, *Child Dev.* 79 (2) (2008) 375–394, <https://doi.org/10.1111/j.1467-8624.2007.01131.x>.
- [8] M. Schneider, R.H. Grabner, J. Paetsch, Mental number line, number line estimation, and mathematical school achievement: their interrelations in Grades 5 and 6, *J. Educ. Psychol.* 101 (2) (2009) 359–372, <https://doi.org/10.1037/a0013840>.
- [9] R.S. Siegler, J.L. Booth, Development of numerical estimation in young children, *Child Dev.* 75 (2) (2004) 428–444, <https://doi.org/10.1111/j.1467-8624.2004.00684.x>.
- [10] C.A. Thompson, R.S. Siegler, Linear numerical-magnitude representations aid children's memory for numbers, *Psychol. Sci.* 21 (9) (2010) 1274–1281, <https://doi.org/10.1177/0956797610378309>.
- [11] L. Yuan, R. Prather, K.S. Mix, L.B. Smith, Number representations drive number-line estimates, *Child Dev.* 91 (4) (2019) e952–e967, <https://doi.org/10.1111/cdev.13333>. Advance online publication.
- [12] J.L. Booth, R.S. Siegler, Developmental and individual differences in pure numerical estimation, *Dev. Psychol.* 42 (1) (2006) 189–201, <https://doi.org/10.1037/0012-1649.41.6.189>.
- [13] J.E. Opfer, C.A. Thompson, D. Kim, Free versus anchored numerical estimation: a unified approach, *Cognition* 149 (2016) 11–17, <https://doi.org/10.1016/j.cognition.2015.11.015>.
- [14] H.C. Barth, A.M. Paladino, The development of numerical estimation: evidence against a representational shift, *Dev. Sci.* 14 (1) (2011) 125–135, <https://doi.org/10.1111/j.1467-7687.2010.00962.x>.
- [15] R.S. Siegler, C.A. Thompson, J.E. Opfer, The logarithmic-to-linear shift: one learning sequence, many tasks, many time scales, *Mind Brain Educ.* 3 (3) (2009) 143–150, <https://doi.org/10.1111/j.1751-228X.2009.01064.x>.
- [16] J.E. Opfer, R.S. Siegler, K.J. Holyoak, R.G. Morrison, *Development of quantitative thinking*. Oxford Handbook of Thinking and Reasoning, Oxford University Press, 2012, pp. 1684–1689.
- [17] R.S. Siegler, G.B. Ramani, Early development of estimation skills, *APS Obs.* 19 (5) (2006) 34–44.
- [18] M.H. Ashcraft, A.M. Moore, Cognitive processes of numerical estimation in children, *J. Exp. Child Psychol.* 111 (2) (2012) 246–267, <https://doi.org/10.1016/j.jecp.2011.08.005>.
- [19] D.J. Cohen, D. Blanc-Goldhammer, P.T. Quinlan, A mathematical model of how people solve most variants of the number-line task, *Cogn. Sci.* 42 (8) (2018) 2621–2647, <https://doi.org/10.1111/cogs.12698>.
- [20] T. Iuculano, B. Butterworth, Rapid Communication: understanding the real value of fractions and decimals, *Q. J. Exp. Psychol.* 64 (11) (2011) 2088–2098, <https://doi.org/10.1080/17470218.2011.604785>.
- [21] R.M. Reinert, M. Hartmann, S. Huber, K. Moeller, Unbounded number line estimation as a measure of numerical estimation, *PLoS One* 14 (3) (2019), e0213102, <https://doi.org/10.1371/journal.pone.0213102>.
- [22] E. Slusser, H. Barth, Intuitive proportion judgment in number-line estimation: converging evidence from multiple tasks, *J. Exp. Child Psychol.* 162 (2017) 181–198, <https://doi.org/10.1016/j.jecp.2017.04.010>.
- [23] D.L. Chesney, P.G. Matthews, Knowledge on the line: manipulating beliefs about the magnitudes of symbolic affects the linearity of line estimation tasks, *Psychon. Bull. Rev.* 20 (6) (2013) 1146–1153, <https://doi.org/10.3758/s13423-013-0446-8>.
- [24] D.J. Cohen, D. Blanc-Goldhammer, Numerical bias in bounded and unbounded number line tasks, *Psychon. Bull. Rev.* 18 (2) (2011) 331–338, <https://doi.org/10.3758/s13423-011-0059-z>.

- [25] D.J. Cohen, P.T. Quinlan, The log-linear response function of the bounded number-line task is unrelated to the psychological representation of quantity, *Psychon. Bull. Rev.* 25 (1) (2018) 447–454, <https://doi.org/10.3758/s13423-017-1290-z>.
- [26] D.J. Cohen, B.W. Sarnecka, Children's number-line estimation shows development of measurement skills (not number representations), *Dev. Psychol.* 50 (6) (2014) 1640–1652, <https://doi.org/10.1037/a0035901>.
- [27] T. Link, H.C. Nuerk, K. Moeller, On the relation between the mental number line and arithmetic competencies, *Q. J. Exp. Psychol.* 67 (8) (2014) 1597–1613, <https://doi.org/10.1080/17470218.2014.892517>.
- [28] J.N. Rouder, D.C. Geary, Children's cognitive representation of the mathematical number line, *Dev. Sci.* 17 (4) (2014) 525–536, <https://doi.org/10.1111/desc.12166>.
- [29] J.L. Sullivan, B.J. Juhasz, T.J. Slattery, H.C. Barth, Adults' number-line estimation strategies: evidence from eye movements, *Psychon. Bull. Rev.* 18 (3) (2011) 557–563, <https://doi.org/10.3758/s13423-011-0081-1>.
- [30] S.L.J. White, D. Szűcs, Representational change and strategy use in children's number line estimation during the first years of primary school, *Behav. Brain Funct.* 8 (1) (2012) 1, <https://doi.org/10.1186/1744-9081-8-1>.
- [31] H.C. Barth, E.B. Slusser, D.J. Cohen, A.M. Paladino, A sense of proportion: commentary on Opfer, Siegler, & Young, *Dev. Sci.* 14 (2011) 1205–1206, <https://doi.org/10.1111/j.1467-7687.2011.01081.x>.
- [32] E.B. Slusser, R.T. Santiago, H.C. Barth, Developmental change in numerical estimation, *J. Exp. Psychol. Gen.* 142 (1) (2013) 193–208, <https://doi.org/10.1037/a0028560>.
- [33] I. Friso-van den Bos, E.H. Kroesbergen, J.E.H. Van Luit, I. Xenidou-Dervou, L. M. Jonkman, M. Van der Schoot, E.C.D.M. Van Lieshout, Longitudinal development of number line estimation and mathematics performance in primary school children, *J. Exp. Child Psychol.* 134 (2015) 12–29, <https://doi.org/10.1016/j.jecp.2015.02.002>.
- [34] M. Hurst, S. Cordes, Rational number comparisons across notation: fractions, decimals, and whole numbers, *J. Exp. Psychol. Hum. Percept. Perform.* 42 (2015) 281–293, <https://doi.org/10.1037/xhp0000140>.
- [35] R.M. Reinert, S. Huber, H.C. Nuerk, K. Moeller, Multiplication facts and the mental number line: evidence from unbounded number line estimation, *Psychol. Res.* 79 (1) (2015) 95–103, <https://doi.org/10.1007/s00426-013-0538-0>.
- [36] R.M. Reinert, S. Huber, H.C. Nuerk, K. Moeller, Strategies in unbounded number line estimation? Evidence from eye-tracking, *Cogn. Process.* 16 (1) (2015) 359–363, <https://doi.org/10.1007/s10339-015-0675-z>.
- [37] R.M. Reinert, S. Huber, H.C. Nuerk, K. Moeller, Sex differences in number line estimation: the role of numerical estimation, *Br. J. Psychol.* 108 (2) (2017) 334–350, <https://doi.org/10.1111/bjop.12203>.
- [38] S. Jung, S. Roesch, E. Klein, T. Dackermann, J. Heller, K. Moeller, The strategy matters: bounded and unbounded number line estimation in secondary school children, *Cogn. Dev.* 53 (2020), 100839, <https://doi.org/10.1016/j.cogdev.2019.100839>.
- [39] R.M. Reinert, K. Moeller, The new unbounded number line estimation task: a systematic literature review, *Acta Psychol.* 219 (2021), 103366, <https://doi.org/10.1016/j.actpsy.2021.103366>.
- [40] R.S. Siegler, Magnitude knowledge: the common core of numerical development, *Dev. Sci.* 19 (3) (2016) 341–361, <https://doi.org/10.1111/desc.12395>.
- [41] R. Cowan, D. Powell, The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability, *J. Educ. Psychol.* 106 (1) (2014) 214–229, <https://doi.org/10.1037/a0034097>.
- [42] V. Gashaj, N. Oberer, F.W. Mast, C.M. Roebers, The relation between executive functions fine motor skills, and basic numerical skills and their relevance for later mathematics achievement, *Early Educ. Dev.* 30 (7) (2019) 913–926, <https://doi.org/10.1080/10409289.2018.1539556>.
- [43] J. Torbeyns, M. Schneider, Z. Xin, R.S. Siegler, Bridging the gap: fraction understandings central to mathematics achievement in students from three different continents, *Learn. Instr.* 37 (2015) 5–13, <https://doi.org/10.1016/j.learninstruc.2014.03.002>.
- [44] D. Sasanguie, B. De Smedt, E. Defever, B. Reynvoet, Association between basic numerical abilities and mathematics achievement, *Br. J. Dev. Psychol.* 30 (2) (2012) 344–357, <https://doi.org/10.1111/j.2044-835X.2011.02048.x>.
- [45] J.L. Booth, R.S. Siegler, Numerical magnitude representations influence arithmetic learning, *Child Dev.* 79 (4) (2008) 1016–1031, <https://doi.org/10.1111/j.1467-8624.2008.01173.x>.
- [46] D.C. Geary, M.K. Hoard, J. Byrd-Craven, L. Nugent, C. Numtee, Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability, *Child Dev.* 78 (4) (2007) 1343–1359, <https://doi.org/10.1111/j.1467-8624.2007.01069.x>.
- [47] D.H. Bailey, R.S. Siegler, D.C. Geary, Early predictors of middle school fraction knowledge, *Dev. Sci.* 17 (5) (2014) 775–785, <https://doi.org/10.1111/desc.12155>.
- [48] L.K. Fazio, D.H. Bailey, C.A. Thompson, R.S. Siegler, Relations of different types of numerical magnitude representations to each other and to mathematics achievement, *J. Exp. Child Psychol.* 123 (2014) 53–72, <https://doi.org/10.1016/j.jecp.2014.01.013>.
- [49] C. Georges, C. Schiltz, Number line tasks and their relation to arithmetics in second to fourth graders, *J. Numer. Cogn.* 7 (1) (2021) 20–41, <https://doi.org/10.5964/jnc.6067>.
- [50] D. Kim, J.E. Opfer, A unified framework for bounded and unbounded numerical estimation, *Dev. Psychol.* 53 (6) (2017) 1088–1097, <https://doi.org/10.1037/dev0000305>.
- [51] J.E. van't Noordende, A.H. van Hoogmoed, W.D. Schot, E.H. Kroesbergen, Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking, *Psychol. Res.* 80 (3) (2016) 368–378, <https://doi.org/10.1007/s00426-015-0736-z>.
- [52] F.A. van der Weijden, E. Kamphorst, R.H. Willemsen, E.H. Kroesbergen, A.H. van Hoogmoed, Strategy use on bounded and unbounded number lines in typically developing adults and adults with dyscalculia: an eye-tracking study, *J. Numer. Cogn.* 4 (2) (2018) 337–359, <https://doi.org/10.5964/jnc.v4i2.115>.
- [53] J. Haffner, K. Baro, P. Parzer, F. Resch, *Heidelberger Rechenstest (HRT): Erfassung Mathematischer Basiskompetenzen im Grundschulalter*, Hogrefe Verlag, 2005.
- [54] J.H. Steiger, Tests of comparing elements of a correlation matrix, *Psychol. Bull.* 87 (2) (1980) 245–251, <https://doi.org/10.1037/0033-2909.87.2.245>.
- [55] T. Dackermann, U. Fischer, S. Huber, H.C. Nuerk, K. Moeller, Training the equidistant principle of number line spacing, *Cogn. Process.* 17 (3) (2016) 243–258, <https://doi.org/10.1007/s10339-016-0763-8>.
- [56] D. Peeters, T. Degrande, M. Ebersbach, L. Verschaffel, K. Luwel, Children's use of number line estimation strategies, *Eur. J. Psychol. Educ.* 31 (2) (2016) 117–134, <https://doi.org/10.1007/s10212-015-0251-z>.
- [57] A.L. Petitto, Development of numberline and measurement concepts, *Cogn. Instr.* 7 (1) (1990) 55–78, https://doi.org/10.1207/s1532690xci0701_3.
- [58] M. Schneider, A. Heine, V. Thaler, J. Torbeyns, B. De Smedt, L. Verschaffel, A. M. Jacobs, E. Stern, A validation of eye movements as a measure of elementary school children's developing number sense, *Cogn. Dev.* 23 (3) (2008) 409–422, <https://doi.org/10.1016/j.cogdev.2008.07.002>.
- [59] P. Bhatia, S. Le Diagon, E. Langlois, M. William, J. Prado, M.L. Gardes, Impact of a game-based intervention on fraction learning for fifth-grade students: a pre-registered randomized controlled study, *J. Comput. Assist. Learn.* 39 (1) (2023) 49–62.
- [60] T. Link, K. Moeller, S. Huber, U. Fischer, H.C. Nuerk, Walk the number line—An embodied training of numerical concepts, *Trends Neurosci. Educ.* 2 (2) (2013) 74–84.