



HAL
open science

The strategy matters: Bounded and unbounded number line estimation in secondary school children

Stefanie Jung, Stephanie Roesch, Elise Klein, Tanja Dackermann, Juergen Heller, Korbinian Moeller

► To cite this version:

Stefanie Jung, Stephanie Roesch, Elise Klein, Tanja Dackermann, Juergen Heller, et al.. The strategy matters: Bounded and unbounded number line estimation in secondary school children. *Cognitive Development*, 2020, 53, pp.100839. 10.1016/j.cogdev.2019.100839 . halshs-02947084

HAL Id: halshs-02947084

<https://shs.hal.science/halshs-02947084>

Submitted on 25 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title:

The strategy matters:

Bounded and unbounded number line estimation in secondary school children

Stefanie Jung^{1,2}, Stephanie Roesch¹, Elise Klein^{1,2}, Tanja Dackermann¹,
Juergen Heller¹, & Korbinian Moeller^{1,2,3}

¹Leibniz-Institut für Wissensmedien,
Schleichstraße 6, 72076 Tuebingen, Germany

²Eberhard Karls University of Tuebingen, Germany
Schleichstraße 4, 72076 Tuebingen, Germany

³LEAD Graduate School and Research Network,
Eberhard Karls University, Tuebingen

Corresponding author:

Dipl. Log. Stefanie Jung

Leibniz-Institut für Wissensmedien

Schleichstraße 6, 72076 Tuebingen, Germany

phone: +49 7071 979-370

email: s.jung@iwm-tuebingen.de

Word count: 8086

ACKNOWLEDGEMENTS

Stefanie Jung was supported by the first Leibniz-WissenschaftsCampus Tuebingen "Bildung und Informationsumwelten" as well as by the German Research Foundation (DFG) providing funding to Elise Klein (KL 2788/2-1). Elise Klein was supported by the Margarete von-Wrangell Fellowship (European Social Fund and the Ministry of Science, Research and the Arts Baden-Württemberg). We thank the participating children, their parents and institutions who made this work possible. We also thank all research assistants for their assistance with data collection.

ABSTRACT:

Changes in number line estimation (NLE) performance are frequently used as an indicator for the development of the number magnitude representation. For this purpose, two different task versions have been applied: a traditional bounded and a relatively new unbounded NLE task. Previous studies mainly assessed primary school children or adults showing that these tasks differ in terms of i) estimation accuracy and solution strategies employed as well as ii) with respect to their relation to other basic numerical/arithmetical skills. So far, data from secondary school children are scarce for bounded NLE, and even no data is available for unbounded NLE. Thus, we assessed bounded and unbounded NLE in grade levels 5 to 7 to evaluate a) developmental as well as strategic influences, and b) the relation of bounded and unbounded NLE performance with basic arithmetic skills. Our results show that children employed the use of different solution strategies for bounded (i.e., proportion-judgement) and unbounded (i.e., magnitude-estimation based) NLE. Moreover, only for bounded NLE estimation accuracy increased with age. Furthermore, estimation performance for bounded but not unbounded NLE was strongly associated with basic arithmetic (addition, subtraction, multiplication, and division). Our findings indicate that the differential results for bounded and unbounded NLE obtained in primary school children seem to generalize to older secondary school children. Presented results substantially contribute to the knowledge about the (consecutive) development of skills pertaining to bounded and unbounded number line estimation.

1 **INTRODUCTION:**
2
3

4 Number line estimation (NLE) refers to the ability to locate numbers on a physical line
5
6 according to their magnitude. NLE performance is a widely used indicator of basic numerical
7
8 skills in studies evaluating the development of numerical skills through childhood. Traditionally,
9
10 NLE refers to estimation performance on a *bounded* number line estimation task. In bounded
11
12 number line estimation, start and endpoint of a specific number range (e.g., 0 to 100) is typically
13
14 specified by a minimum value indicated at the left end of the number line and a maximum
15
16 value at the right end of the line. Participants are required to estimate the spatial position of a
17
18 given number (e.g., Siegler and Opfer, 2003). Compared to that, *unbounded* NLE performance
19
20 has been assessed less frequently, but more and more recently (e.g., Cohen & Blanc-
21
22 Goldhammer, 2011; Link, Huber, Nuerk, & Moeller, 2014; Reinert, Hartmann, Huber, & Moeller,
23
24 2019). In the unbounded NLE task, the number line has no labelled endpoint, but a single unit
25
26 distance (e.g., 0-1) in addition to the start point. Starting from this unit distance, target numbers
27
28 have to be located upon the given line.
29
30
31
32
33
34
35
36

37
38
39 Bounded and unbounded NLE tasks have been used as indicators to draw conclusions from
40
41 task performance onto the underlying magnitude representation and its development (e.g., for
42
43 bounded NLE: Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Booth & Siegler, 2006,
44
45 2008; Ebersbach, Luwel, & Verschaffel, 2015; Siegler & Booth, 2004; Siegler & Opfer, 2003;
46
47 Thompson & Opfer, 2010; Young & Booth, 2015; e.g., for unbounded NLE: Cohen & Blanc-
48
49 Goldhammer, 2011; Link, Huber, et al., 2014). However, the two tasks were found to differ in
50
51 terms of i) the solution strategies employed (e.g., considering the usage of reference points [cf.
52
53 Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Link, Huber, et al., 2014]) and ii)
54
55 their association with other basic numerical as well as arithmetic skills (for a meta-analysis see
56
57
58
59
60
61
62
63
64
65

1 Schneider et al., 2018). More importantly, despite numerous studies relating primarily bounded
2
3 NLE to other basic numerical and arithmetic skills, the vast majority of these studies focused
4
5 either on young children from kindergarten to the end of primary school (e.g., Link, Nuerk, et
6
7 al., 2014; Moeller et al., 2009; Schneider et al., 2017; Zhu et al., 2017) or adults (Reinert, Huber,
8
9 Nuerk, & Moeller, 2017, 2015). So far, the age range in-between (i.e., secondary school children)
10
11 has largely been neglected for bounded NLE. For unbounded NLE, there is still no empirical
12
13 data at all from secondary school children. Crucially, consideration of relatively narrow age
14
15 ranges strongly limits general statements on the development of number representation.
16
17
18
19
20
21

22 In the current study, we aimed at addressing this gap by investigating bounded and
23
24 unbounded NLE in secondary school children¹ (i.e., 5th to 7th graders) to evaluate both a)
25
26 developmental changes in *estimation accuracy and strategy use* and b) the relation of bounded
27
28 and unbounded *number line estimation with basic arithmetic*.
29
30
31

32 33 34 *Estimation accuracy and strategy use*

35
36
37 Regarding bounded NLE, there are different views on the developmental changes of NLE
38
39 performance and its relationship to the underlying representation of number magnitude.
40
41 Generally, all views are based on the observation that children of different age groups perform
42
43 differently in the same number range: Younger children (i.e., preschoolers and first-graders)
44
45 were observed to systematically overestimate the spatial position of small numbers (i.e., placing
46
47 9 at about the location where 40 would go on a 0-100 scale; Moeller et al., 2009). In turn, they
48
49 localised larger numbers in a compressed way towards the upper bound of the scale (e.g.,
50
51 Siegler & Booth, 2004; Siegler & Opfer, 2003). Older children (e.g., second-graders), in contrast,
52
53
54
55
56
57
58
59
60

61 ¹ In the German school system, majority of secondary schools begins with the 5th grade.
62
63
64
65

1 did not show such a pattern and were observed to perform more accurately in the same
2
3
4 number range (e.g., Siegler & Booth, 2004).
5

6
7 These estimation patterns have been discussed in various theoretical accounts (see
8
9 Dackermann et al., 2015, for a recent review). Most importantly, two competing accounts can
10
11 be differentiated: the log-to-linear shift account (e.g., Booth & Siegler, 2006; Siegler & Booth,
12
13 2004; Siegler & Opfer, 2003) and the proportion-judgement account (e.g., Barth & Paladino,
14
15 2011; Slusser, Santiago, & Barth, 2013).
16
17
18
19
20

21 The log-to-linear shift account proposes a representational change from an initially logarithmic
22
23 layout, as indicated by overestimation of smaller numbers, to a more linear layout of estimation
24
25 patterns with increasing age and experience. Accordingly, a logarithmic function was found to
26
27 fit the observed estimation pattern best in younger children whereas a linear function reflected
28
29 the observed estimation patterns best in older children and adults (Booth & Siegler, 2006, 2008;
30
31 Siegler & Booth, 2004; Siegler & Opfer, 2003). The account assumes estimation patterns to
32
33 directly reflect the underlying magnitude representation (Siegler & Opfer, 2003). However, the
34
35 interpretation of a direct assessment of the magnitude representation is questioned in
36
37 alternative accounts that emphasize the importance of other numerical processes in addition
38
39 to the pure magnitude estimation. Ebersbach, Luwel, Frick, Onghena, and Verschaffel (2008),
40
41 for instance, mentioned the increasing familiarity of children with numbers (but see Thompson
42
43 & Opfer, 2010 for results contradicting the assumption that children's familiarity with numbers
44
45 predicts individual's mental representation of the number line), whereas Moeller et al., 2009
46
47 suggested to consider the understanding of the place-value system, which also affects NLE
48
49 performance.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 The most prominent alternative account, the proportion-judgement account, postulates that
2
3 potential reference points (i.e., start and end point of the respective scale, but also quartiles
4
5 and midpoint, also referred to as [subjective] landmarks) are considered for estimation (e.g.,
6
7 Barth & Paladino, 2011; Siegler & Opfer, 2003; Slusser et al., 2013). The use of reference points
8
9 is supposed to increase with age and to enhance estimation accuracy (e.g., Slusser et al., 2013;
10
11 see also Schneider et al., 2008, for eye-fixation data.). Evidence for the account comes from
12
13 children's estimation patterns: while target estimates seem to follow a linear layout, estimation
14
15 errors for target numbers were more accurate and varied less at and/or around reference
16
17 points, which results in a typical *M*-shaped pattern in contour analyses (Ashcraft & Moore,
18
19 2012; Cohen & Blanc-Goldhammer, 2011; Link, Huber, et al., 2014). Additionally, use of
20
21 reference points and thus proportion-judgement strategies were found to be indexed by (one-
22
23 or two-) cyclic power models (Barth & Paladino, 2011; Hollands & Dyre, 2000). Contrary, fit of
24
25 cyclic power models was also associated with feedback on the mid-point of the number line
26
27 (i.e., 50 by a number line length of 100) provided to participants (Opfer, Siegler, & Young, 2011;
28
29 Opfer, Thompson, & Kim, 2016).

30
31
32 In sum, all accounts share the idea that several different development-related cognitive skills
33
34 and solution strategies are involved in bounded NLE (cf. Barth & Paladino, 2011; Cohen &
35
36 Sarnecka, 2014; Ebersbach et al., 2008; Moeller et al., 2009; Siegler & Booth, 2004; Siegler &
37
38 Opfer, 2003; Slusser et al., 2013). This is reflected by an ongoing debate on the suitability of
39
40 the bounded NLE task as a valid measure of children's number magnitude representation (e.g.,
41
42 Cohen & Sarnecka, 2014; Slusser & Barth, 2017), because processes of magnitude estimation
43
44 seem to be masked by other cognitive processes and solution strategies involved in performing
45
46 the task.
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 In contrast, unbounded NLE is argued to reflect a purer measure of number magnitude
2
3 representation (e.g., Cohen & Blanc-Goldhammer, 2011; Link, Huber, et al., 2014; see also
4
5 Reinert, Huber, Nuerk, & Moeller, 2015 for eye-tracking evidence supporting this argument)
6
7 and seems to be less influenced by age as no development-related performance changes have
8
9 been found (Cohen & Sarnecka, 2014; Link, Huber, et al., 2014). This assumption is based on
10
11 empirical observations of children of different age levels within primary school and adults. Error
12
13 variability in unbounded NLE was found to increase linearly with number magnitude (Cohen &
14
15 Blanc-Goldhammer, 2011; Link, Huber, et al., 2014). This finding was interpreted to reflect the
16
17 strategic procedure of estimating a multiple of the given unit distance (i.e., dead-reckoning
18
19 strategy as suggested by Cohen & Blanc-Goldhammer, 2011 or a counting-like strategy as
20
21 recommended by Schneider et al., 2018), which leads to an accumulation of estimation errors
22
23 (Reinert et al., 2015). As such, it was argued that unbounded NLE reflects a more direct type of
24
25 number magnitude estimation (Link, Huber, et al., 2014, Reinert et al., 2019; but see Kim &
26
27 Opfer, 2017). Cohen & Blanc-Goldhammer (2011) showed first evidence that dead-reckoning
28
29 strategy, for instance, is indicated by a repetitive scalloped estimation pattern.
30
31
32
33
34
35
36
37
38
39
40

41 *Number line estimation and basic arithmetic*

42
43 Numerous studies have examined associations of bounded and/or unbounded NLE and basic
44
45 arithmetic. These studies have shown significant correlations, but only for the bounded version
46
47 of the task (e.g., Link, Huber, et al., 2014; Schneider et al., 2018). Bounded NLE performance
48
49 significantly predicts mathematical achievement concurrently but also longitudinally (e.g.,
50
51 Booth & Siegler, 2008; Schneider, Grabner, & Paetsch, 2009; but see Simms et al., 2016, for
52
53 methodological aspects) and correlates with counting, basic arithmetic and algebra as well as
54
55 with school achievement and grades (Schneider et al., 2018, for a meta-analysis) from
56
57
58
59
60
61
62
63
64
65

1 kindergarten to primary school (Booth & Siegler, 2006). In this vein, Friso-van den Bos et al.
2
3
4 (2015) argued that these associations may occur due to a mutual relation between both basic
5
6 arithmetic and NLE skills in numerical development: With increasing practice in solving
7
8 mathematical problems in childhood, the representation of number magnitude becomes more
9
10 accurate, which is in turn reflected in better NLE performance.
11
12

13
14 Comparing data from both versions of the NLE task, Cohen and Sarnecka (Cohen & Sarnecka,
15
16 2014) argued that bounded and unbounded NLE should to be associated with different basic
17
18 arithmetic skills. The authors found that young children (i.e., preschoolers) underperformed in
19
20 bounded NLE, but were able to cope with the unbounded version approximately as well as
21
22 older children. Cohen and Sarnecka (2014) concluded that unbounded NLE may be less
23
24 demanding, probably because task solution requires less advanced basic arithmetic skills (i.e.,
25
26 addition and multiplication) for using dead-reckoning strategies as compared to bounded NLE
27
28 (i.e., subtraction and division) for applying proportion judgement strategies.
29
30
31
32
33

34
35
36 Contrasting bounded and unbounded NLE performance in a sample of fourth-graders, the
37
38 study of Link, Nuerk and Moeller (2014) partially confirmed these theoretical assumptions. The
39
40 authors argued that bounded NLE may require a division of the number line to guide the
41
42 application of proportion-based strategies by reference points. Subsequently, starting from a
43
44 reference point, a leftward (reflecting subtraction) or rightward shift on the number line
45
46 (reflecting addition) might be necessary to locate the respective target number correctly upon
47
48 the line. This adjustment, however, is not appropriate in unbounded NLE. In this vein, Link and
49
50 colleagues (2014) did not observe significant correlations between basic arithmetic and
51
52 unbounded NLE in their sample of fourth-graders (but see Reinert et al., 2015 for different
53
54 results regarding multiplication). Despite an increasing number of studies directly comparing
55
56
57
58
59
60
61
62
63
64
65

1 bounded and unbounded NLE, it is not yet clear how both tasks differ in their associations with
2
3 various basic arithmetic operations.
4
5

6
7 However, it is important that the results and assumptions described above are again mainly
8
9 based on observations of children from kindergarten to the end of primary school and of
10
11 adults, whereas the age group in-between, i.e., secondary school children, has so far largely
12
13 been neglected. Crucially, the few studies investigating the relationship between NLE and basic
14
15 arithmetic in typically developing secondary school children mainly assessed NLE of fractions
16
17 and related estimation performance to either a standardised mathematical school achievement
18
19 test (Siegler & Pyke, 2013) or to school grades (Schneider et al., 2009). Only few studies focused
20
21 on whole numbers. Some of these studies included rather small sample sizes (32 sixth graders
22
23 in Siegler & Opfer, 2003 and 24 sixth graders Thompson & Opfer, 2010) or did not use standard
24
25 NLE tasks by evaluating NLE on both negative (-10,000-0) and positive numbers (-1,000-1,000)
26
27 in sixth and seventh graders (Young & Booth, 2015).
28
29
30
31
32
33
34
35

36 On a neuro-functional level, Berteletti et al. (2015) provided first evidence that bounded NLE
37
38 performance correlates with activation of brain areas associated with number magnitude
39
40 processing (such as the intraparietal sulcus during single digit subtraction). Nevertheless,
41
42 empirical data from secondary school children on unbounded NLE are still missing entirely.
43
44 Moreover, so far, there is no comparison of bounded and unbounded NLE in secondary school
45
46 children, which also compares the association of bounded and unbounded NLE performance
47
48 and basic arithmetic in three consecutive age levels.
49
50
51
52
53

54 **The current study**

55
56
57
58
59 The current study focused on bounded and unbounded NLE in secondary school children from
60
61 grade 5 to 7. In particular, we investigated whether bounded NLE performance changes
62
63
64
65

1 between these age groups and aimed at providing first empirical data on how secondary school
2
3 children perform in unbounded NLE. For bounded NLE, previous studies assessing number line
4
5 estimation beyond primary school used number ranges between 0-1,000 and 0-100,000 (Link,
6
7 Huber, et al., 2014; Siegler & Opfer, 2003; Thompson & Opfer, 2010; Young & Booth for
8
9 negative numbers, 2015). For unbounded NLE, mainly smaller number ranges up to 20 (Cohen
10
11 & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014), 29 (Link, Nuerk, et al., 2014; Link, Huber,
12
13 et al., 2014), and 50 (Reinert et al., 2015a; Reinert et al., 2015b, Reinert et al., 2019) were used.
14
15 Only Kim and Opfer (2017) used number ranges of 0-30/100/1,000 in unbounded NLE.
16
17 However, Cohen, Blanc-Goldhammer, and Quinlan (2018) pointed to the actual task design
18
19 used by Kim and Opfer (2017): in unbounded NLE, children and adults typically overestimate
20
21 target numbers (see also Cohen and Sarnecka, 2014). Hence, the physical number line needs
22
23 to leave enough room beyond the largest target number so that participants can demonstrate
24
25 not only this bias but also scalar variance in their estimations. This issue demonstrates that it is
26
27 impossible to accurately run an unbounded number-line task with large numbers (cf. Kim and
28
29 Opfer, 2017).
30
31

32
33 However, in studies directly comparing bounded and unbounded NLE performance either
34
35 equal number ranges (Reinert, Huber, Nuerk, & Moeller, 2015, Cohen & Blanc-Goldhammer,
36
37 2011; Cohen & Sarnecka, 2014) or different number ranges were applied in both tasks (e.g.,
38
39 from 0-100, 0-1,000, and 0-10,000 in bounded NLE and 0-25 in unbounded NLE; Link, Huber,
40
41 Nuerk, & Moeller, 2014, Reinert et al., 2018, or 0-49 in unbounded NLE; Reinert et al., 2019).
42
43 Crucially, Link and colleagues (2014) provided evidence that participants performed less
44
45 accurate in unbounded NLE even though the number range was very small in the unbounded
46
47 (i.e., 0-25) as compared to the number range in the bounded NLE task (i.e., 0-10,000).
48
49 Considering above described problems with creating an unbounded number line for large
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 target values and to tie on previous results, we used the same number ranges as in Link et al.,
2
3
4 (2014) study. Importantly, and different from bounded number line estimation, there do not
5
6 seem to be substantial influences of age in unbounded NLE (e.g., Link et al., 2014). In fact,
7
8 estimation patterns for all age groups were characterised by estimation errors increasing in
9
10 size and in variability with increasing magnitude of the respective target numbers.
11
12

13
14 Furthermore, to address the ongoing debate on which version of the NLE task – bounded or
15
16 unbounded – is associated with which basic arithmetic operation, the relation of both bounded
17
18 and unbounded NLE with the four basic arithmetic operations (i.e. addition, subtraction,
19
20 multiplication, and division) was examined. Based on the findings from previous studies
21
22 outlined above, our hypotheses were as follows:
23
24
25
26

27
28 First, we expected to find significant differences between bounded and unbounded NLE (cf.
29
30 Cohen & Sarnecka, 2014). In particular, mean estimation errors were assumed to be higher in
31
32 the unbounded task version (Link, Huber, et al., 2014). Second, regarding developmental
33
34 changes, we expected a decline of estimation errors with age for bounded NLE, but no
35
36 significant age-related differences in estimation errors for unbounded NLE (Cohen & Blanc-
37
38 Goldhammer, 2011; Link, Huber, et al., 2014). Third, we expected children to use proportion-
39
40 judgement strategies in bounded but not in unbounded NLE and thus, only bounded NLE
41
42 errors to follow a characteristic *M*-shaped distribution pattern (cf. Ashcraft & Moore, 2012). The
43
44 use of potential reference points in bounded NLE should be predominantly reflected by cyclic
45
46 models that best fit the estimation pattern of children. Finally, based on the theoretical
47
48 considerations of Cohen and Sarnecka (2014) and the findings of Link et al. (2014), we assumed
49
50 both tasks to be associated with different basic arithmetic skills. In particular, we supposed NLE
51
52 performance in the bounded task to be associated with subtraction, addition, and division.
53
54
55
56
57
58
59
60
61
62
63
64
65

1 However, we did not formulate a well-directed hypothesis for the association of unbounded
2
3 NLE performance and basic arithmetic. Results of the present study may help to further clarify
4
5 these relations.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **METHODS:**
2
3

4 ***Participants:***
5
6
7

8 A total of 989 children (463 girls) from 12 German secondary schools participated in this study:
9
10 children were recruited from 17 5th grade classes (n=342, mean age: 10.58, SD=8.24 months),
11
12 21 6th grade classes (n=383, mean age: 11.67, SD=8.48 months) and 15 7th grade classes
13
14 (n=264, mean age: 12.57, SD=8.19 months). Cross-sectional data was collected at two different
15
16 survey points (SP1 and SP2). Children were divided into two samples and completed a
17
18 computerised assessment of either a bounded (Sample 1 assessed at SP1 and SP2: n=700 [SP1:
19
20 n=389 and SP2: n=311] mean age: 11.53, SD=12.47 months) or unbounded (Sample 2 assessed
21
22 only at SP2: n=289 mean age: 11.56, SD=12.93 months) NLE task. Each child was examined
23
24 once. Furthermore, children's basic arithmetic as well as their spelling skills were assessed as
25
26 control variables.
27
28
29
30
31
32

33
34 Data from 195 children who did not complete either the spelling or the arithmetic tasks (n=94)
35
36 or had more than 33% missing data (n= 101) in the NLE tasks after outlier correction were not
37
38 considered for further analyses. Finally, complete data of 573 children for the bounded NLE
39
40 task and 221 children for the unbounded NLE task entered the analyses.
41
42
43
44

45
46 Written informed consent was obtained from parents prior to the study besides children's
47
48 verbal assent before actual testing. The study was approved by the local ethics committee (LEK
49
50 2014/19) as well as by the school authority.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **Tasks and procedure:**
2
3

4 Data collection took place during regular classes and lasted about 45 minutes. Arithmetic skills
5 were assessed separately for each basic arithmetic operation (i.e., addition, subtraction,
6 multiplication, division) together with either the bounded or unbounded NLE task. Spelling was
7 examined with respect to most relevant German spelling rules (i.e., capitalization, consonant
8 doubling, lengthening, rule words) and served as control variable. Children completed the
9 arithmetic and spelling tasks in counterbalanced order across participants. All stimuli were
10 presented on a 15.6-inch Lenovo ThinkPad T530 laptop display with a resolution of 1024 x 768
11 pixels at normal viewing distance and with target items in black against a white background.
12
13
14
15
16
17
18
19
20
21
22
23
24
25

26 Task relevant laptop functions (e.g., capitalization, deletion of incorrect entries) were
27 introduced immediately before the assessment, followed by at least one practice trial for each
28 task. Instructions were presented both visually and auditory. This procedure ensured that all
29 children had the same prerequisites to cope with the tasks in the computerised assessment.
30
31
32
33
34
35
36

37 **Basic arithmetic**
38
39

40 Children's basic arithmetic skills in *addition*, *subtraction*, *multiplication*, and *division* were
41 assessed following this particular order using a production paradigm. Children were asked to
42 type in the correct result of a presented problem. Overall, 20 problems were presented for each
43 operation. Stimuli were selected as follows: in *addition* (e.g., $36+47=$ ___) and *subtraction* (e.g.,
44 $91-67=$ ___), problems were controlled for problem size for each operation. In particular, in
45 *addition*, four problems each had a result of the 6th to the 9th decade, two problems from the
46 5th decade and one problem each from the 2nd to the 4th decade. Half of the problems
47 required a carry operation. In *subtraction*, ten problems with a result in the range 10-19 were
48 chosen. The results of the other ten problems fell in the range from 20 to 29. Half of the
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 problems required a borrowing operation. In *multiplication* (e.g., $8*7=$ ___), three problems
2
3 each were picked from the tables of 3 to 9 as well as two problems from the table of 5. Neither
4
5 ties nor problems in reversed order were presented. For *division* (e.g., $56/7=$ ___), multiplication
6
7 problems were reversed. Table A1.1 in Appendix A1 provides an overview of all arithmetic
8
9 problems.
10
11

12 *Number line estimation*

13
14
15 Children completed either a bounded (range 0-10,000) or an unbounded (range 0-29) version
16
17 of the NLE task. In the bounded and unbounded NLE task, target items were presented on the
18
19 left end above the lower bound of the number line. The physical length of the number line was
20
21 716 pixels. Children were required to mark the correct position of a given number by means of
22
23 the mouse cursor. Number line and target number remained on the screen until a response
24
25 was given by a mouse click. Both NLE tasks were preceded by a practice trial (5,000 for the
26
27 bounded NLE task and 1 for the unbounded NLE task²) to ensure that the children understood
28
29 the task and to familiarize the children with the use of the computer mouse. In the practice trial
30
31 no feedback was provided on the children's estimation accuracy. In the bounded version (range
32
33 0-10,000), a total of 24 target numbers were presented in randomised order (i.e., 74, 135, 1097,
34
35 1203, 2137, 2315, 3408, 3476, 4542, 4712, 4798, 4957, 5103, 5239, 5298, 5372, 6594, 6781, 7685,
36
37 7812, 8793, 8946, 9786, and 9851). In the unbounded version (range 0-29), participants had to
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 ² Please note that the choice of practice items can be discussed: for bounded NLE, there is an ongoing
53 debate on how anchoring and targeted practice on the midpoint of the number line is associated with
54 increased estimation accuracy (Slusser, Santiago, & Barth, 2013; Opfer, Thompson, & Kim, 2016).
55 However, Dackermann and colleagues (2018) did not observe a reliable effect of explicit midpoint
56 instruction. In the present study, children located the number serving as practice item (i.e., 5000) on the
57 number line but did not receive any feedback as to whether their estimate was correct or that the given
58 number represented the mid-point of the number line. For unbounded NLE, we used the smallest unit
59 size in the practice trial to ensure that children understand to use (multiples of) the unit size for their
60 estimation. Again, children did not receive any feedback on their estimate.
61
62
63
64
65

1 position 24 numbers on a number line, of which only the start and a unit (i.e., the length
2
3
4 corresponding to 1) but no end-point were indicated. The set of target numbers to be located
5
6 ranged from 2 to 25, leaving sufficient distance between the end of the physical line (i.e., 29)
7
8 and the largest target number (i.e., 25) as children tend to overestimate in this task.
9

10 11 *Spelling competencies*

12
13
14
15
16 Spelling assessment followed a writing-to-dictation task which required children to complete
17
18 gapped sentences (e.g., "Eine Bananenschale ist ..." [Abfall]; engl. "A banana peel is ..."
19
20 [garbage]). First, each sentence was presented auditorily, followed by a repetition of the target
21
22 word (depicted in square brackets). Children were instructed to read along the sentences
23
24 during dictation, wait until the target word was repeated and then type in the respective target
25
26 word. Overall, 55 test words were administered with respect to the following orthographic
27
28 rules: capitalization (55 words, e.g., '**A**bfall', [engl. 'garbage']), consonant doubling (21 words,
29
30 e.g., 'Ab**f**all', [engl. 'garbage']), lengthening (14 words, e.g., '**Die**b', [engl., 'thief']) and rule words
31
32 (i.e., words for which the spelling cannot be derived by the phoneme structure; 14 words, e.g.,
33
34 '**L**ok', [engl. 'loco']). In this example, the short vowel is not followed by a double consonant, as
35
36 one would assume from the German pronunciation). Table A1.2 in Appendix A1 presents the
37
38 entire test material for the spelling assessment. Individual spelling was assessed automatically
39
40 as either correct (scored 1) or incorrect (scored 0) with respect to each orthographic rule in
41
42 each test word (e.g., '**A**bfall' [engl. 'garbage'] was assessed concerning (inner sentence)
43
44 capitalization and consonant doubling; correct spelling of '**A**bfall' leads to a word score of max.
45
46 2) resulting in a maximum overall sum score of 104.
47
48
49
50
51
52
53
54
55
56
57

58 **Data Analysis:**

59 60 61 *Estimation accuracy and strategy use*

1 First, to increase comparability of the results, we standardised children's estimation errors by
2
3 the length of the respective number line in both tasks. Therefore, children's general estimation
4
5 accuracy was analyzed using the percent absolute error as dependent variable ($PAE = \frac{|Estimate - Target\ number|}{Scale} * 100$; cf. Siegler and Booth, 2004). PAE is the standard dependent
6
7
8
9
10
11 variable in the literature when evaluating NLE performance across different number ranges,
12
13 age groups, versions of number line estimation tasks, etc. (e.g., Booth & Siegler, 2008; Link et.,
14
15 2014; Siegler & Booth, 2004; Slusser et al., 2012; Slusser & Barth, 2017; Young & Booth, 2015).
16
17
18 Consider the following example for the bounded NLE task with the target number 74. When
19
20
21 the child locates the number at the position of 234, he or she is 160 units off. The PAE in this
22
23
24 example amounts to 1.6% ($PAE = \frac{|234 - 74|}{10,000} * 100$). As PAE does not allow to indicate
25
26
27 the pattern of over- and underestimation, we also calculated the mean percent relative
28
29
30 estimation error [$REE = \frac{(Estimate - Target\ number)}{Scale} * 100$], which would be +1.6%
31
32
33 indicating slight overestimation. For the REE, a value of zero means an accurate estimate. A
34
35
36 negative value reflects underestimation, whereas a positive value indicates overestimation.
37
38
39 Furthermore, contour analyses (cf. Ashcraft & Moore, 2012) were conducted separately for each
40
41
42 grade to contrast children's estimation errors around and in between potential reference points
43
44
45 (e.g. start, end, quartile, and mid points) as an index of the use of proportion-judgement
46
47
48 strategies (cf. Ashcraft & Moore, 2012) in bounded NLE. Estimates of the following numbers
49
50
51 were pooled to reflect the usage of potential reference points beginning from start (74 and
52
53
54 135) via midpoint (4957 and 5103) to the end point (9851 and 9786). Numbers around the first
55
56
57 (2315 and 3408) and third quartile (6781 and 7685) in-between these reference points were
58
59
60 combined (cf. Lai Zax & Barth, 2018; for other pooling approaches) In unbounded NLE, use of
61
62
63 reference points was not expected as the absence of an end point of the scale prevents the
64
65
66 possibility of a reference. Nevertheless, for better comparability of the results, estimates of

1 target numbers around the start (2 and 3), first quartile (7 and 8) mid (13 and 14), third quartile
2
3
4 (19 and 20) and end point (24 and 25) of the number line scale were combined as well. Data
5
6 points of the contour analysis were then compared using an analysis of variance (ANOVA).
7
8

9
10 Additionally, we conducted model fittings on children`s individual estimates to validate number
11
12 line estimation strategies derived from contour analyses. Data were modeled with Matlab 9.4.0
13
14 (MATLAB and Statistics Toolbox Release R2018a, The MathWorks, Inc., Natick, Massachusetts,
15
16 United States) using the trust region algorithm for non-linear model fitting.
17
18
19
20

21 *Number line estimation and basic arithmetic*

22
23

24 Association of number line estimation performance and basic arithmetic operations was
25
26 assessed by means of partial correlations. With respect to correlation analyses, the Shapiro–
27
28 Wilk test (Shapiro & Wilk, 1965) was conducted in order to assess normal distribution. As
29
30 distribution of basic arithmetic and spelling data deviated significantly from normal
31
32 distribution, Spearman correlation coefficients were calculated. In particular, semi-partial rank
33
34 correlation coefficients of non-parametric data ("Spearman's rho") using the R package ppcor
35
36 (Kim, 2015) were calculated to examine the correlation of general accuracy in number line
37
38 estimation and in the four basic arithmetic operations (i.e., addition, subtraction, multiplication,
39
40 division), controlling for age and spelling ability. In order to validate results, cross-validation
41
42 methods were applied. For this purpose, existing samples were randomly divided into two data
43
44 sets each. As cross-validation revealed reliable outcomes in line with results for the overall
45
46 sample, only latter results are presented below. Cross-validation results are reported in Table
47
48 A2 in Appendix A2. Bonferroni-Holm procedure (Holm, 1979) was applied in order to
49
50 (adequately) correct for multiple testing.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

RESULTS:

In total, complete data sets of 573 children for the bounded NLE task, in particular, 198 fifth graders (94 female, mean age: 10.48, $SD = 7.32$ months), 231 sixth-graders (105 female, mean age: 11.69, $SD = 8.64$ months) and 144 seventh-graders (74 female, mean age: 12.54, $SD = 8.28$) entered analyses. For unbounded NLE, 221 data sets were analyzed, this means, data of 86 fifth graders (41 female, mean age: 10.62, $SD = 8.05$ months), 81 sixth-graders (37 female, mean age: 11.72 $SD = 8.28$ months) and 54 seventh-graders (22 female, mean age: 12.75, $SD = 9.30$ months). Table 1 provides descriptive information regarding all variables used in the study.

Table 1:

A) Demographics bounded sample

	Items (n)	Fifth grade		Sixth grade		Seventh grade	
		Mean	SD	Mean	SD	Mean	SD
Spelling (sum score)	104	83.44	11.71	85.55	11.84	91.59	8.65
Addition	20	18.24	2.79	17.94	3.03	18.59	2.64
Subtraction	20	14.98	4.85	14.11	5.43	15.93	4.39
Multiplication	20	17.47	3.38	17.21	3.84	18.08	2.30
Division	20	17.90	2.97	17.75	3.16	18.52	1.82
NLE task mean PAE	24	5.94	1.66	5.49	1.28	5.17	1.42

B) Demographics unbounded sample

	Items (n)	Fifth grade		Sixth grade		Seventh grade	
		Mean	SD	Mean	SD	Mean	SD
Spelling (sum score)	104	83.46	11.95	87.25	11.44	90.74	9.23
Addition	20	18.49	2.28	18.14	2.73	19.24	1.70
Subtraction	20	15.22	4.73	14.35	4.90	16.43	4.18
Multiplication	20	17.37	3.34	18.20	2.18	18.18	2.37
Division	20	17.90	3.46	18.54	2.48	18.28	2.51
NLE task mean PAE	24	9.77	4.16	9.76	4.39	8.97	3.36

Note: In the spelling assessment, children could achieve a maximum overall score of 104 points; in arithmetic assessment, children could achieve a maximum score of 20 points for each operation (i.e., addition, subtraction, multiplication, division).

1 Children of the two samples did not differ significantly according to their sum scores in terms
2
3 of spelling (5th grade: $t(158.82) = -0.018, p = 0.99$, 6th grade: $t(144.31) = -1.13, p = 0.26$, 7th
4
5 grade: $t(90.12) = 0.59, p = 0.59$) and basic arithmetic (5th grade: $t(160.86) = -0.26, p = 0.79$, 6th
6
7 grade: $t(197.5) = -1.71, p = 0.09$, 7th grade: $t(100.4) = -0.72, p = 0.47$) as assessed with Welch
8
9 two sample t -tests. Moreover, at SP₂ bounded and unbounded NLE performance was assessed
10
11 within the same school classes. Thus, school- and classroom related influences (e.g., school
12
13 location, class size, teaching style, average math performance) were equal for both groups.
14
15
16
17
18

19
20 A 2 (condition: bounded vs. unbounded NLET) x 3 (5th vs. 6th vs. 7th grade) analysis of variance
21
22 (ANOVA) on PAEs revealed a significant main effect for condition [$F(1,788) = 412.21, p < .001$,
23
24 $\eta^2_p = 0.35$], such that average PAE was significantly lower for bounded NLE ($M = 5.57, SD =$
25
26 1.49) than for unbounded NLE ($M = 9.57, SD = 4.07$). The main effect for grade [$F(2,788) = 7.95$,
27
28 $p < .001, \eta^2_p = 0.02$] indicated a significant influence of grade levels on NLE performance.
29
30 Furthermore, there was no interaction between condition and group [$F(2,788) = 0.605, p > 0.05$,
31
32 $\eta^2_p < 0.01$]. Post hoc comparisons for grade using Tukey HSD test applying an adjusted p -level
33
34 for multiple comparisons showed a marginally significant [$p = .056$] difference between mean
35
36 PAE of children in the 5th ($M=5.95, SD=1.66$) and 7th grade ($M=5.17, SD=1.42$), in the bounded
37
38 condition. In the unbounded condition, no significant differences in mean PAEs were observed
39
40 between grade levels [all $p > .284$; 5-graders ($M=9.77, SD=4.16$), 6-graders ($M=9.76, SD=4.39$),
41
42 and 7-graders ($M=8.97, SD=3.36$)].
43
44
45
46
47
48
49
50
51
52
53
54

55 *Number line estimation accuracy and strategy use*

56
57
58

59 Figure 1 reflects children's number line estimation patterns for both the bounded (Panel A) and
60
61 the unbounded NLE task (Panel B) separated for grade levels.
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

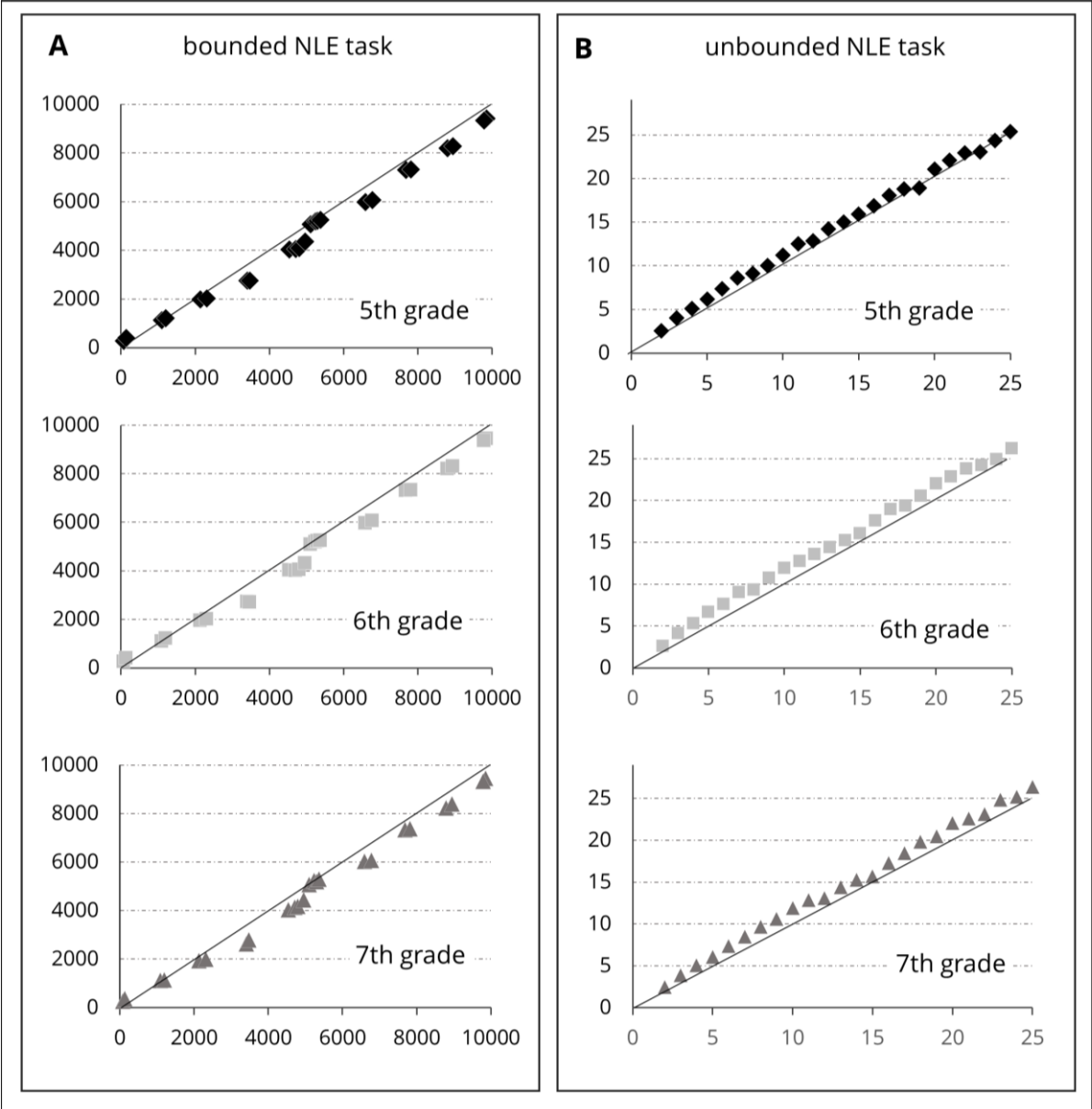


Figure 1: Mean estimation pattern for bounded (A) and unbounded (B) number line estimation: The figure shows target values (x-axis) and mean estimates (y-axis) for all three age levels, with $x=y$ in the diagonal.

Crucially, Figure 1 reveals that estimation patterns differ between bounded and unbounded NLE: Figure 2 depicts the mean REE score for both bounded and unbounded NLE. Most importantly, children underestimated target numbers in the bounded NLE task, whereas they overestimated target numbers in the unbounded NLE task.

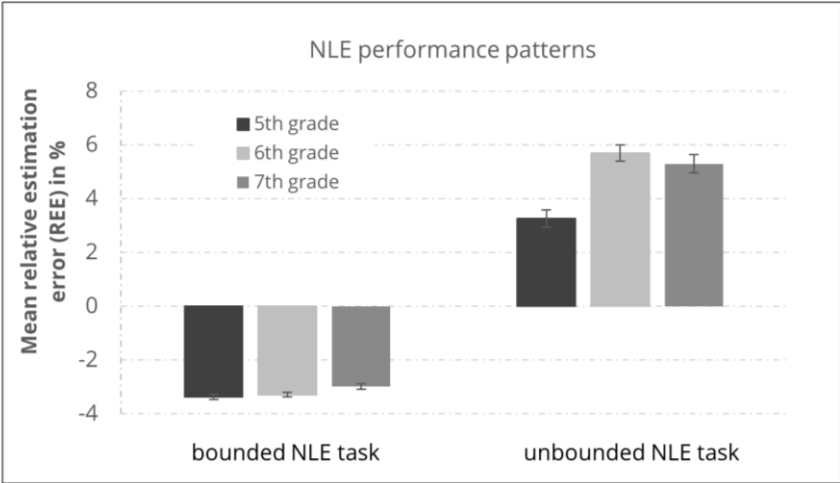


Figure 2: Mean relative estimation errors (REE) for bounded and unbounded number line estimation performance.

Regarding children’s performance in the bounded NLE task Figure 3 revealed estimation errors to vary for different target numbers. In contrast, children’s estimation errors increased linearly in the unbounded NLE task. Summarizing estimation errors at/around reference points, contour analyses revealed an *M*-shaped pattern of error distribution (see Figure 3A) which has been argued to be characteristic for proportion-judgement strategies (cf. Cohen and Blanc-Goldhammer, 2011). Correlating PAEs and size of target number revealed no significant correlations for all grade levels (all $r < 0.333$, all $p > 0.11$).

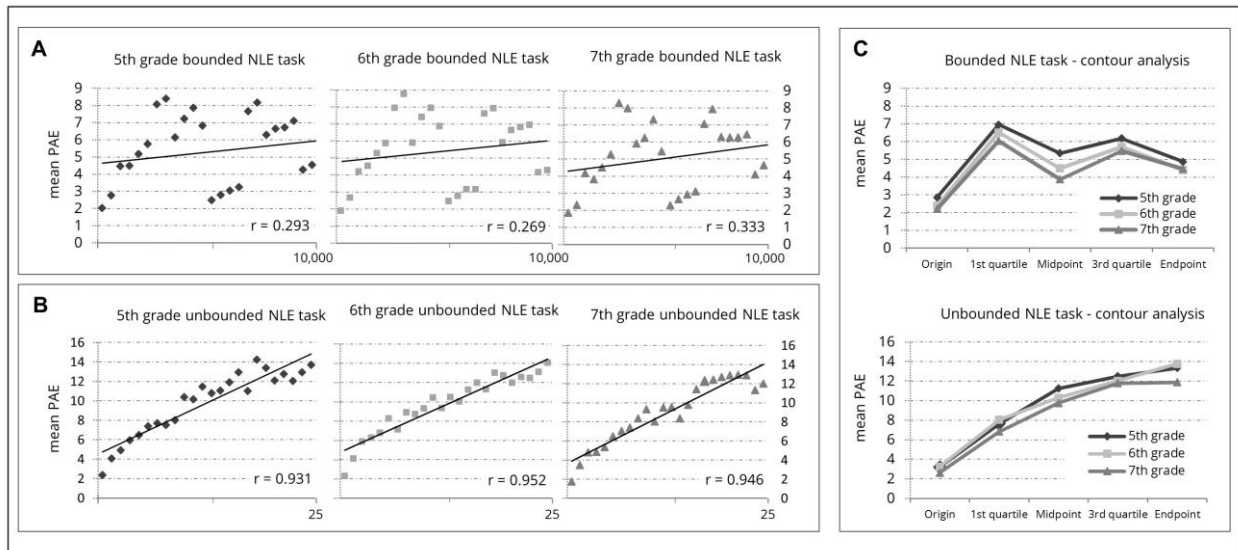


Figure 3: Mean percentage absolute errors (PAE) for bounded and unbounded number line estimation performance. The left panel depicts mean PAE and target numbers in the (A) bounded and (B) unbounded NLE task separately for all grade levels. The right panel (C) shows contour analyses summarizing percent absolute errors (PAE) at specific reference points (cf. Ashcraft and Moore, 2012) for all grade levels (i.e., grade 5-7) again for bounded (top) and unbounded (bottom) NLE.

Bonferroni-Holm corrected *t*-tests were conducted to compare PAEs at specific reference points (start-, mid- and endpoint) and in the intervals between them (first and third quartile). Results substantiated the descriptive pattern of the contour analyses by indicating significantly reduced PAEs at reference points as compared to the intervals in-between for all grades [5th grade: $t(192) > -7.23, p < .001$, Cohen's *d* = 0.23 – 1.14; 6th grade: $t(227) > -21.53, p < .001$, Cohen's *d* = 0.40 – 1.43, and 7th grade: $t(142) > -14.79, p < .001$, Cohen's *d* = 0.37 – 1.23]. In unbounded NLE, in contrast, linearly increasing PAEs across all age groups were observed: PAEs increased monotonically with target number (see Figure 3B) resulting in significant correlations between PAEs and size of target number (from $r = 0.931$ to $r = 0.952$, all $p < 0.05$). This pattern was also found in contour analyses (Figure 3C). Bonferroni-Holm corrected *t*-tests revealed significant differences between both the start- and midpoint and the quartile [5th grade: $t(85) > -12.09, p < .001$, Cohen's *d* = 0.58 – 1.30; 6th grade: $t(80) > -12.09, p < .001$, Cohen's *d* = 0.25 – 1.50, and 7th grade: $t(53) > -10.14, p < .001$, Cohen's *d* = 0.31 – 1.38], but not for the third quartile and the endpoint [5th grade: $t(85) > -1.64, p > 0.05$, Cohen's *d* = 0.18; and 7th grade:

1 $t(53) > -2.25, p(\text{adj.}) > 0.05$, Cohen's $d = 0.01$), except for the 6th grade ($t(80) > 2.5432, p < .001$,
2
3
4 Cohen's $d = 0.28$].
5
6

7 Additionally, we fitted children's number line estimates with different mathematical models.
8
9 Table 2 shows absolute and relative frequencies (in parenthesis) of best fitting models for
10 bounded and unbounded NLE separated for the three different grade levels.
11
12
13
14
15

16 In bounded NLE, linear models provided best fit for all grade levels (for 56% of fifth, 55% of
17 sixth, and 56% of seventh graders). Only about 30% of all children's estimates were best
18 determined by models indicating proportion-judgement strategies (i.e., one-cyclic or two-
19 cyclic models), and thus the use of reference points. This finding may contradict evidence for
20 proportion-judgement from the contour analyses.
21
22
23
24
25
26
27
28

29 In unbounded NLE, over 90% of children's estimates were accounted best by linear (for 63% of
30 the fifth, 68% of the sixth, and 65% of the seventh graders) or power models (for 29% of the
31 fifth, 30% of the sixth, and 33% of the seventh graders) for all grade levels. Model fitting did
32 not indicate model fits in favor of dead-reckoning in unbounded number line estimation.
33
34
35
36
37
38
39
40 Crucially, and in particular for unbounded NLE, mean adjusted R^2 resp. the AIC values were
41 almost identical. As such, a unique derivation of the best fitting model seems not feasible.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 2: Model fittings for bounded and unbounded number line estimation

A	Bounded NLE task				B	Unbounded NLE task			
		5th	6th	7th			5th	6th	7th
	N	198	232	144		N	86	81	54
linear model	adj R ²	0.945	0.954	0.962		adj R ²	0.877	0.884	0.891
	AIC	281.453	283.290	287.868	linear model	AIC	35.573	37.565	37.684
	Δ_AIC	0	0	0		Δ_AIC	0	0	0
	n	112 (56)	128 (55)	81 (56)		n	54 (63)	55 (68)	35 (65)
adj R ²	0.944	0.954	0.961			adj R ²	0.874	0.878	0.883
Power model	AIC	281.655	283.569	288.116	power model	AIC	36.599	38.917	39.385
	Δ_AIC	0.20	0.28	0.25		Δ_AIC	1.026	1.352	1.702
	n	32 (16)	27 (12)	20 (14)		n	25 (29)	24 (30)	18 (33)
	adj R ²	0.929	0.936	0.945			adj R ²	0.876	0.879
one-cyclic model	AIC	287.959	290.357	295.015	dual scallop model	AIC	36.734	39.512	39.785
	Δ_AIC	6.51	7.07	7.15		Δ_AIC	1.16	1.95	2.10
	N	31 (16)	35 (15)	23 (16)		n	6 (7)	2 (2)	1 (2)
	adj R ²	0.9273	0.9388	0.9472			adj R ²	0.872	0.878
two-cyclic model	AIC	288.749	290.450	295.241	mutli scallop model	AIC	37.300	39.520	39.697
	Δ_AIC	7.30	7.16	7.37		Δ_AIC	1.73	1.96	2.01
	N	23 (12)	41 (18)	20 (14)		n	1 (1)	0 (0)	0 (0)

Note: Absolute and relative frequency (percentages) of best fitting models for bounded (A) and unbounded (B) number line estimation separated for grade levels. The best fitting models are indicated in bold script.

Regarding the partially contradictory results of contour analyses and model fittings for bounded NLE, closer visual examination of children's individual estimation pattern was more revealing. Table 3 provides results of individual contour analysis for each child separated for the best fitting models as well as for grade level. The table differentiates between the M-shape pattern of number estimates and other patterns (i.e., N-shape, inverse N-shape, V-shape, etc.). At first glance, more than one third of fifths and sixth graders showed an M-shape estimation pattern on the individual level. This number increased in the seventh grade up to almost fifty percent indicating an increasing use of reference points with age. Yet, these findings cannot be clearly derived from the results of individual model fittings: successful use of reference points (i.e., proportion-judgement) and thus minimal estimation errors can be also accounted by linear models (e.g., Link, Huber, et al., 2014). Contrary, unsuccessful use of reference points (i.e., a V-shape or N-shape pattern) can be accounted by one- or two-cyclic models.

Table 3: Individual contour analyses and model fittings for bounded NLE

	5th grade		6th grade		7th grade	
	M-Shape	Other	M-Shape	Other	M-Shape	Other
linear model	41	71	50	78	40	41
power model	8	24	10	17	8	12
one-cyclic model	8	23	14	21	13	10
two-cyclic model:	6	17	9	32	6	14
Σ	63 (32)	135 (68)	83 (36)	148 (64)	67 (47)	77 (53)

Note: Absolute and relative frequency (percentages) for the use of proportion-judgement strategy (i.e., M-Shape estimation pattern) or other strategies derived from contour analysis in bounded number line estimation.

Number line estimation and basic arithmetic

First, we analyzed to what extent spelling and arithmetic skills were associated to control for influences of general cognitive abilities in subsequent analyses. Partial correlations controlling for age revealed relatively high correlations between basic arithmetic skills and overall spelling skills for both samples assessed with the bounded ($n=573$, $r=.51$, $p<.001$) and unbounded ($n=221$, $r = .43$, $p<.001$) NLE task. These correlations indicated a strong association between spelling and arithmetic performance. As spelling and arithmetic skills have been found to rely on common general cognitive abilities (e.g., working memory, verbal memory, and reasoning skills, e.g., Knievel, Daseking, & Petermann, 2010; Passolunghi, Mammarella, & Altoè, 2008), we used overall spelling skills as a proxy to control for influences of general cognitive ability. Subsequently, partial correlations were calculated controlling for age and spelling skills. Table 4 shows the results of the partial correlation between number line estimation performance and basic arithmetic separately for grade levels. For bounded NLE, strength of correlations increased with age. In grade 5, NLE performance correlated significantly with subtraction (i.e., lower error rate in subtraction problems was associated with less estimation errors in NLE), but not with addition, multiplication, or division. In grade 6, significant correlations were found for addition, subtraction and multiplication, but not for division. In grade 7, bounded NLE

performance was found to be significantly associated with all basic arithmetic operations, but highest with addition and subtraction. For unbounded NLE, no significant correlation was found between task performance and basic arithmetic.

Table 4: Partial correlations (one-tailed) between PAEs of bounded (left panel) and unbounded NLE and basic arithmetic

<i>Grade</i>	Bounded NLE task				Unbounded NLE task			
	<i>Add</i>	<i>Sub</i>	<i>Mul</i>	<i>Div</i>	<i>Add</i>	<i>Sub</i>	<i>Mul</i>	<i>Div</i>
5 th grade	-0.09	-0.20**	-0.08	-0.03	0.06	-0.14	0.03	-0.06
6 th grade	-0.29**	-0.23**	-0.21**	-0.06	-0.08	-0.05	0.02	0.06
7 th grade	-0.43**	-0.44**	-0.24**	-0.25*	0.03	-0.17	0.14	0.05

Note: Overall partial correlations were calculated for all grade levels. Add=Addition, Sub=Subtraction, Mul=Multiplication, Div=Division. * $p < .05$, ** $p < .001$.

1 **DISCUSSION:**
2
3

4 In the following, we discuss results on number line *estimation accuracy and strategy use* and
5 the association of *NLE and basic arithmetic* separately with regard to the relevant literature.
6
7

8
9
10 *Estimation accuracy and strategy use*
11

12
13 Results of the current study substantiated that secondary school children performed differently
14 in bounded and unbounded NLE: PAEs were almost twice as high in unbounded than in
15 bounded NLE. These findings are consistent with previous results on primary school children
16 (i.e., grade 3 and 4) and adults (e.g., Link, Huber, et al., 2014, Reinert et al., 2019; for a production
17 version of the task). Importantly, this result cannot be explained by the different number ranges
18 used in the present study for unbounded and bounded NLE. Although one may assume that
19 all children were familiar with both intervals, the larger number interval should nevertheless be
20 more error prone due to the problem size effect (Zbrodoff & Logan, 2005, for a review).
21
22 However, mean estimation errors in bounded NLE were significantly lower. More likely, the
23 difference in estimation accuracy arises from different solutions strategies applied. In bounded
24 NLE, application of proportion-based strategies simplifies task solution; the presence of a start
25 and an end point facilitated the derivation of the center of the number line and also
26 decomposition into sections. This procedure reduces the number of mistakes made by the
27 children. Thus, estimation errors were reduced. Unbounded NLE does not offer this solution
28 strategy as the end point is missing.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 Contour analysis in the present study substantiated this assumption. In line with Ashcraft and
54 Moore (2012) as well as with Link et al. (2014), a characteristic *M*-shaped pattern of estimation
55 errors (i.e., smaller and less variable estimation errors for target numbers at or around reference
56 points) was observed in contour analyses for bounded NLE. This suggested that children used
57
58
59
60
61
62
63
64
65

1 proportion-based strategies to solve the task. By contrast, we observed an almost linearly
2
3 increasing error pattern for unbounded NLE, which has been previously assumed to be related
4
5 to a magnitude estimation-based strategy in unbounded NLE (Cohen & Blanc-Goldhammer,
6
7 2011; Link, Huber, et al., 2014). However, a closer look at the potential reference in the
8
9 unbounded NLE tasks disclosed that the children may not have been able to express their
10
11 entire estimation bias (with overestimating target numbers by about 50%, cf. Cohen &
12
13 Sarnecka, 2014) because of the length of the physical number line (i.e., 716 pixels
14
15 corresponding to 29 units). This might explain the lack of significant further increase in
16
17 estimation error for target numbers in the 3rd quartile and at around the end point (i.e., target
18
19 numbers larger than 20). However, we are confident that this does not affect general
20
21 conclusions made about solution strategies in unbounded NLE. Nevertheless, future studies
22
23 should be aware of the necessity to provide a sufficiently long physical number line (depending
24
25 on the number ranged used) to be able to assess the respective (over)estimation bias.
26
27

28
29 However, these findings were only partially corroborated by model fitting results. Model
30
31 fittings for bounded NLE indicated best fit of linear models for the majority of children's
32
33 estimates. As recently discussed by Link and colleagues (2014), a very accurate estimation
34
35 pattern can hardly be differentiated by model fittings. Contrarily, model fitting results for
36
37 unbounded NLE were unambiguous. They did not provide any evidence for specific strategy
38
39 use (i.e., proportion-judgement) in unbounded number line estimation.
40
41
42
43
44
45
46
47
48
49
50

51
52 Furthermore, as with primary school children (e.g., Barth & Paladino, 2011; Cohen & Blanc-
53
54 Goldhammer, 2011; Link, Huber, et al., 2014; Slusser & Barth, 2017), estimation accuracy in the
55
56 present study was found to improve with age for bounded NLE. In contrast, for unbounded
57
58 NLE no significant differences in estimation accuracy was observed across grade levels.
59
60
61
62
63
64
65

1 However, we observed age-related increases from 5th to 7th grade for both basic arithmetic
2 and spelling performance which were considered as control variables. Considering the
3 observation that estimation performance does not change in unbounded NLE across the
4 different age groups, the question arises to what extent unbounded NLE performance is
5 influenced by general age-related developmental processes.
6
7
8
9
10
11
12

13 *Number line estimation and basic arithmetic*

14
15
16
17
18 Results substantiated an association of NLE performance for bounded but not unbounded NLE
19 and basic arithmetic (cf. Cohen & Sarnecka, 2014; Link et al., 2014). Correlation coefficients
20 were comparable to those reported in other studies (ranging from $r = 0.29 - 0.86$) relating
21 various mathematical competencies (i.e., problem solving, calculation; see Schneider et al.,
22 2018, for an overview). In particular, we observed subtraction skills to be associated with
23 bounded NLE performance consistently for all grades. In grade 6, correlations with addition
24 and multiplication were of equal strength as with subtraction. In grade 7, all four arithmetic
25 operations were significantly associated with bounded NLE.
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 We suggest that children who are well versed in solving subtraction tasks may be more likely
41 to successfully use proportion-judgement strategies. The following factors might contribute to
42 this finding: First, addition and subtraction are taught as the first arithmetic operations in
43 school (e.g., Bildungsplan Sekundarstufe I, Ministerium für Kultus, Jugend und Sport, Baden
44 Württemberg, 2016). However, subtraction is regarded to be more demanding as compared to
45 addition (e.g., because of borrowing, Artemenko et al., 2018), and children were found to use
46 calculation (i.e., instead of fact retrieval in simple addition and multiplication problems) to solve
47 subtraction problems even in 2-digit numbers (e.g., Ischebeck et al., 2006). Thus, subtraction
48 might be more sensitive to reflect differences in magnitude estimation skills as it also relies on
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 magnitude manipulations to a higher degree. Second, according to the proportion-judgement
2
3
4 account, a leftward (reflecting subtraction) or rightward shift (reflecting addition) starting from
5
6 a reference point might be necessary to locate the number correctly upon the line (cf. Cohen
7
8 and Sarnecka, 2014). A closer descriptive look at the results of the contour analysis (Figure 3)
9
10 and mean PAEs at the quartiles supports this assumption: starting from the mid-point, a
11
12 leftward shift (from mid-point to the first quartile reflecting subtraction) resulted in higher
13
14 estimation errors than a shift to the right (from mid-point to the third quartile reflecting
15
16 addition). Third, further evidence may be derived from neuro-functional level: Berteletti and
17
18 colleagues (2015) demonstrated that activation of brain areas associated with number
19
20 magnitude processing (as required in subtraction) were correlated with NLE performance.
21
22 Furthermore, Ischebeck et al. (2006) provided neuro-functional evidence that a training of
23
24 subtraction problems in young adults encouraged the application of efficient procedural
25
26 solution strategies. This result suggests that children with high subtraction skills are more likely
27
28 to apply sufficient solution strategies (i.e., proportion-judgement) at least in bounded NLE.
29
30
31
32
33
34
35
36
37

38 An explanation for significant associations between bounded NLE and multiplication as well as
39
40 division occurring not before grade 6 and 7, respectively, might be that, according to German
41
42 math curricula, multiplication and division are increasingly used with the introduction of
43
44 fractions in grade 6 (e.g., Bildungsplan Sekundarstufe I, Ministerium für Kultus, Jugend und
45
46 Sport, Baden Württemberg, 2016). Although all four basic arithmetic operations have already
47
48 been acquired in primary school (i.e., before grade 5; Huber, Moeller, & Nuerk, 2012), the
49
50 increasing significance of multiplication and division (i.e., fractions) from grade 6 onwards
51
52 might enable children to apply proportion-judgement strategies more proficiently, leading to
53
54 better performance in bounded NLE. Visual inspection of contours on individual levels
55
56 confirmed the increasing use of reference points with age. This in turn can influence NLE
57
58
59
60
61
62
63
64
65

1 performance as presented by Siegler and colleagues (2011) in another domain of mathematics
2
3 (i.e., fractions) showing that segmentation strategies involving dividing up the number line with
4
5 subjective landmarks leads to more accurate fraction number line estimates.
6
7

8
9 Moreover, correlation coefficients became stronger for subtraction, addition, multiplication
10
11 and finally division from grade 5 to 7, suggesting an increasing importance of basic arithmetic
12
13 for bounded NLE. These findings substantiate the assumption of Friso-van den Bos et al. (2015)
14
15 who argued that associations between arithmetic skills and NLE are predominantly driven by
16
17 developmental processes.
18
19
20
21

22
23 For the association of unbounded NLE performance and basic arithmetic, we did not find
24
25 significant relations with any basic arithmetic operation in secondary school children (Link et
26
27 al., 2014; for similar results). Building on the results of Cohen and Blanc-Goldhammer (2011)
28
29 and Reinert et al. (2015), we would have expected at least a weak correlation of unbounded
30
31 NLE with addition and multiplication, respectively. Even though, and unlike Reinert et al. (2015),
32
33 we did neither focus on analyzing different unit distances (0-1, 0-5, etc.) nor specifically
34
35 selected multiples of these as target numbers. However, the lack of associations between
36
37 unbounded NLE and basic arithmetic seems plausible considering the assumption that
38
39 unbounded NLE may be solved by iterative counting like strategies (Reinert et al., 2015;
40
41 Schneider et al., 2018). Counting as a highly automated process may not depend on basic
42
43 numerical and arithmetical skills. In this vein, unbounded NLE seems to reflect purer numerical
44
45 estimation, which may be superimposed less by underlying arithmetic processes (i.e., addition,
46
47 subtraction, division) of other solution strategies as is proportion-judgement (see also Cohen
48
49 & Blanc-Goldhammer, 2011; Link, Huber, et al., 2014; Reinert et al., 2015). In sum, the present
50
51 data provide additional evidence from secondary school children that unbounded number line
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 estimation may indeed be solved primarily based on processing and estimating number
2
3
4 magnitude.

5
6
7 When interpreting these results, there are some constraints that need to be considered. First,
8
9 in the present study different number ranges were used in bounded (0-10,000) and unbounded
10
11 (0-29) number line estimation to tie on results of previous studies (e.g., Link, Huber, et al., 2014;
12
13 Siegler & Opfer, 2003; Thompson & Opfer, 2010). As introduced above, there are sensible
14
15 reasons for using larger number ranges for bounded NLE and to use small number ranges in
16
17 unbounded NLE (e.g., Cohen et al., 2018). However, to enhance comparability of results,
18
19 standardised measures considering the length of the respective number lines were used (see
20
21 also Link, Huber, et al., 2014; for a methodological discussion on the application of different
22
23 number ranges). This approach seems warranted as we did not draw any conclusions on
24
25 differences in task difficulty between differently large number ranges.
26
27
28
29
30
31
32

33
34 Second, the study was cross-sectional, so we did not monitor intra-individual development in
35
36 bounded and unbounded NLE. However, conclusions on numerical development with age
37
38 seem warranted as for both spelling and basic arithmetic tasks our data documented increasing
39
40 performance with increasing age. Nevertheless, it would be desirable to pursue intra-individual
41
42 developmental changes in estimation accuracy in secondary school children more
43
44 systematically in future studies.
45
46
47
48
49

50
51 Third, as we assessed performance in bounded and unbounded NLE inter-individually, direct
52
53 comparisons of both NLE tasks within one child were not possible. Nevertheless, both samples
54
55 seem well comparable as bounded and unbounded NLE was assessed within the same school
56
57 classes (at least at SP₂), and thus, class-room related influences (e.g., class size, teaching style,
58
59 average math performance of the school class) were equally distributed across both groups.
60
61
62
63
64
65

1 Moreover, and this is particularly important, the two groups solving either bounded or
2
3 unbounded NLE neither differed significantly in mean basic arithmetic skills nor in spelling
4
5 performance. We therefore assume that both groups are well comparable and sufficient to
6
7 draw comparable conclusions on bounded and unbounded NLE performance. However,
8
9 individual comparisons would be desirable for future studies. Finally, it is not clear yet whether
10
11 and how the presentation medium (i.e., computerised assessment in the current study)
12
13 moderates NLE (Schneider et al., 2018), and potentially leads to inconsistent results (e.g.,
14
15 increased error variability) as compared to earlier paper-pencil studies. Future mode effect
16
17 studies on NLE comparing computerised and paper-pencil testing could provide further
18
19 insights.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **CONCLUSIONS:**

2
3
4 In the current study, we investigated bounded and unbounded NLE in secondary school
5 children (i.e., grade 5 to 7). We found that children’s performance in bounded NLE was
6 significantly better as compared to unbounded NLE for all grade levels. Moreover, estimation
7 accuracy seems to improve with age only in the bounded, but not in the unbounded task
8 version. With respect to the association of NLE performance with basic arithmetic, we observed
9 that estimation performance in bounded but not unbounded NLE was strongly associated with
10 basic arithmetic operations (i.e., addition, subtraction, multiplication, division). Interestingly,
11 these associations increased with age, and thus, are probably subject to developmental
12 changes. With respect to solution strategies employed in NLE, current results corroborated the
13 assumption of bounded but not unbounded NLE to rely on proportion-based estimation
14 strategies.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32
33
34 In summary, these data indicate that differential results on the association between bounded
35 and unbounded NLE with arithmetic operations obtained in primary school generalize to older
36 secondary school children. Thereby, our results substantially contribute to the knowledge
37 about the (consecutive) development of skills pertaining to bounded and unbounded number
38 line estimation, as they provide additional evidence for the assumption that unbounded NLE
39 may indeed reflect a more direct and pure measure of number magnitude estimation.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

APPENDIX A1

Appendix A1 provides test material used to assess basic arithmetic (Table A1.1) and spelling competencies (Table A1.2) of this study.

Table A1.1: Basic arithmetic problems

N	Addition	Subtraction	Multi- plication	Division
P	$12 + 3 = 15$			
1	$36+47 = 83$	$91-67 = 24$	$8 \times 7 = 56$	$48 \div 6 = 8$
2	$29+67 = 96$	$93-67 = 26$	$8 \times 6 = 48$	$56 \div 7 = 8$
3	$25+67 = 92$	$51-37 = 14$	$8 \times 4 = 32$	$32 \div 4 = 8$
4	$21+27 = 48$	$94-69 = 25$	$7 \times 6 = 42$	$54 \div 9 = 6$
5	$65+28 = 93$	$92-76 = 16$	$9 \times 7 = 63$	$42 \div 6 = 7$
6	$29+38 = 67$	$81-65 = 16$	$6 \times 9 = 54$	$63 \div 7 = 9$
7	$38+25 = 63$	$41-18 = 23$	$4 \times 7 = 28$	$24 \div 8 = 3$
8	$27+29 = 56$	$45-29 = 16$	$4 \times 6 = 24$	$72 \div 8 = 9$
9	$24+68 = 92$	$52-27 = 25$	$4 \times 9 = 36$	$35 \div 5 = 7$
10	$46+36 = 82$	$33-18 = 15$	$3 \times 8 = 24$	$21 \div 3 = 7$
11	$53+36 = 89$	$96-72 = 24$	$9 \times 8 = 72$	$18 \div 3 = 6$
12	$42+34 = 76$	$97-71 = 26$	$3 \times 9 = 27$	$36 \div 9 = 4$
13	$65+21 = 86$	$36-14 = 22$	$5 \times 8 = 40$	$28 \div 7 = 4$
14	$28+51 = 79$	$76-62 = 14$	$3 \times 4 = 12$	$40 \div 8 = 5$
15	$37+41 = 78$	$39-26 = 13$	$7 \times 5 = 35$	$24 \div 6 = 4$
16	$24+45 = 69$	$29-16 = 13$	$6 \times 3 = 18$	$27 \div 9 = 3$
17	$41+35 = 76$	$56-41 = 15$	$7 \times 3 = 21$	$12 \div 4 = 3$
18	$15+13 = 28$	$38-21 = 17$	$6 \times 5 = 30$	$30 \div 5 = 6$
19	$13+24 = 37$	$39-11 = 28$	$5 \times 4 = 20$	$20 \div 4 = 5$
20	$41+13 = 54$	$27-13 = 14$	$5 \times 3 = 15$	$15 \div 3 = 5$

Note: Basic arithmetic problems were preceded by one practice trial (P) to familiarize the children with the computerised testing.

Table A1.2: Spelling competencies assessed by gapped sentences

N	Target	C	D	L	R	Gapped sentence	English translation
P	Fuß	x			x	Lisa trat mir auf meinen rechten ...	Lisa stepped on my right ... (foot)
1	Abfall	x	x			Eine Bananenschale ist...	A banana peel is ... (garbage)
2	abmessen	x	x			Mit dem Lineal kann man Längen..	Use the ruler to ... lengths. (measure)
3	Baby	x			x	In der Wiege schläft ein...	In the cradle sleeps a ... (baby)
4	beißen	x				Hunde, die bellen, ... nicht.	Dogs that bark don't ... (bite)
5	Beschmutzen	x	x			Im Park ist das... der Bänke verboten.	It is forbidden to ... the benches in the park. (stain)
6	bieten	x		x		Will man eine Auktion gewinnen, muss man hoch...	If you want to win an auction, you have to ... high. (bid)
7	bitten	x	x			Wenn die Gäste klingeln,... wir sie herein.	We' ll ... them in when the guests ring the bell. (invite)
8	Blumenwiese	x		x		Das Mädchen pflückte Blumen von der ...	The girl picked flowers from the ... (field)
9	Bus	x			x	Beeil dich, wir kommen zu spät zum...	Hurry up, we'll be late for the ... (bus)
10	Chili	x			x	Nimm für das Essen bitte wenig...	Please take a little ... for your meal. (chili)
11	cool	x			x	Seinen neuen Haarschnitt findet Hannes richtig...	Hannes thinks his new haircut is really ... (cool)
12	dehnen	x		x		Vor dem Sport sollte man die Muskeln...	Before doing sports you should ... your muscles. (stretch)
13	dick	x	x			Ein Buch mit 800 Seiten ist...	An 800-page book is ... (thick)
14	Dieb	x		x		Die Handtasche wurde von einem... geklaut.	The purse was stolen by a ... (thief)
15	Entlassung	x	x			Es ist traurig, wenn einem Arbeiter mit der... gedroht wird.	It is sad when a worker is faced with a ...
16	erfahren	x		x		Die Nachricht ist unglaublich, Sarah muss unbedingt davon...	The news is incredible, Sarah needs to ... about it. (know)
17	erzählst	x		x		Die Geschichten sind immer spannend, wenn du sie...	The stories are always exciting when you ... them. (tell)
18	Eselohr	x		x		Eine umgeknickte Seite eines Buches nennt man...	A folded page of a book is called a... (dog's ear)
19	fällt	x		x		Es ist schon kalt und die Temperatur ... ständig noch weiter.	It's already cold and the temperature keeps ... (falling)
20	Fenstergriff	x	x			Jedes Fenster muss einen ... haben.	Every window must have a window ... (handle)
21	Geburtstag	x				Meist gibt es einen Kuchen zum...	Usually there is a cake for your ... (birthday)
22	grässlich	x	x			Das verbrannte Essen schmeckte...	The burnt food tasted ... (awful)
23	heißen	x	x			Manche finden das Kunstwerk schön, andere finden es...	Some people like the work of art, others consider it to be ... (ugly)
24	hell	x	x			Die Sonne scheint ...	The sun is shining ... (bright)
25	knurrend	x	x			Ich traute mich nicht ins Haus, weil ein Hund... vor der Türe saß.	I didn't dare come in the house because a dog was ... at the door. (grumbling)
26	Lok	x			x	Die Waggonen werden von der... gezogen	The wagons are pulled by the ... (loco)
27	Lot	x			x	Die Mauer muss nach dem... ausgerichtet werden.	The wall must be aligned to the ... (perpendicular)

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

28	Target	x		Gapped sentence	English translation
29	mahlt	x	x x	Die Mühlsteine drehen sich, wenn der Müller das Korn...	The millstones turn when the miller ... the grain. (grinds)
30	malt	x		Oma freut sich, wenn Lisa ein Bild für sie...	Grandma is happy when Lisa ... a picture for her. (draws).
31	Mandarinen	x	x	Zu Weihnachten gibt es viele Nüsse und...	There are lots of nuts and ... for Christmas. (tangerines)
32	Maß	x	x	Zentimeter sind ein ... für Längen.	Centimeters are a ... of length. (measure)
33	Mus	x	x	Die reifen Äpfel kochen wir zu...	We cook the ripe apples to ... (puree)
34	Nachsitzen	x	x	Wenn man zu viel Quatsch macht, drohen manche Lehrer mit...	Some teachers threaten ... if you do too much rubbish. (detention)
35	näher	x	x	Von Italien nach Frankreich ist es... als vom Mond zur Erde.	From Italy to France it is ... than from the moon to the earth. (closer)
36	okay	x	x	Den neuen Kinofilm findet Max ganz...	Max thinks the new movie is (okay)
37	Orange	x	x	In den Obstsalat gehört auch eine...	An ... also belongs in the fruit salad. (orange)
38	Pony	x	x	Auf der Weide steht ein kleines...	There's a little ... in the pasture. (pony)
39	Portion	x	x	Im Sommer isst man gerne eine große...Eis.	In summer you like to eat a large ... of ice cream. (portion)
40	Pralinen	x	x	Oma liebt Süßes, am meisten mag sie...	Grandma loves sweets, she likes ... the most. (pralines)
41	Rechnen	x		Mit Bruchzahlen ist das... schwieriger als mit ganzen Zahlen.	Fractions are more difficult to ... than integers. (calculate)
42	schafft	x	x	Die Fans feuern den Fahrer an, damit er es auf den ersten Platz ...	All fans cheer on the racer to ... it to the first place. (make)
43	schmierig	x	x	Wenn man zu viel Gel nimmt, werden die Haare oft ...	If you take too much gel, the hair often becomes ... (greasy).
44	sehnt	x	x	Wenn dich die Arbeit anstrengt,... du dich nach Pausen	When you work hard, you ... for breaks. (long)
45	Straßenbahn	x	x	Ohne Ticket darf man nicht mit der ... fahren	You may not take the ... without a ticket. (tram)
46	stumm	x	x	Ich hatte einen Schock, konnte nichts sagen und blieb ganz ...	I was in shock, couldn't say anything and remained completely ... (mute).
47	stützt	x	x	Der Kranke ... sich auf die Krücke	The patient ... on the crutch. (leans)
48	treffen	x	x	Können wir uns in der Pause auf dem Hof ...	Can we ... on the yard during the break. (meet)
49	Üben	x		Ohne ... wird keiner ein Meister	Without ..., no one will be a master. (practice)
50	Überschwemmung	x	x	Wenn ein Fluss über die Ufer tritt, gibt es eine...	When a river bursts its banks, there is a ... (flood)
51	verspritzen	x	x	Beim Patronenwechsel bin ich vorsichtig, ich will die Tinte nicht...	I'm careful when changing cartridges; I don't want to ... the ink. (splash)
52	verwöhnt	x	x	Meine Schwester ist ein Nesthäkchen, sie wird meistens...	My sister is a nestling, she is usually ... (pampered)
53	voll	x	x	Schenke mir mein Glas bitte ganz...	Give me a ... glass, please. (full)
54	Zaubertrick	x	x	Wer ein Kaninchen aus einem Hut holt, kann einen ...	When you take a rabbit out of a hat, you can do a ... (magic trick)
55	Zwerg	x		Im Märchen taucht oft ein kleiner... auf.	A little ... often appears in fairy tales. (dwarf)

Note: Test words and corresponding sentences are presented in alphabetic order. The table also specifies which German spelling rules appear in which test word (C= capitalization, D= consonant doubling, L=lengthening, R=rule words). The English translation is intended to give the interested non-German-speaking reader an impression of the used test material. However, it is not suitable to reflect the particular characteristics of the German orthographic rules for the test words in the spelling assessment.

1 **APPENDIX A2**

2
3
4 In order to validate results of partial correlation analyses, cross-validation methods were
5
6 applied. For this purpose, existing samples were randomly divided into two data sets (n_1 and
7
8 n_2) for the bounded and unbounded NLE task.
9

10
11
12 First, associations of spelling and basic arithmetic in the overall sample were compared to the
13
14 two sub samples. Partial correlations controlling for age revealed relatively high correlations
15
16 between basic arithmetic skills and spelling skills in both samples assessing bounded ($n=573$,
17
18 $r=.51$, $p<.001$) and unbounded ($n=221$, $r = .43$, $p<.001$) NLE performance. Cross-validated
19
20 correlations were found to be similarly highly correlated (bounded NLET: $n_1=287$, $r_1=.43$,
21
22 $p_1<.001$, $n_2=286$, $r_2=.49$, $p_2<.001$; unbounded NLET: $n_1=111$, $r_1=.41$, $p_1<.001$, $n_2=110$, $r_2=.42$,
23
24 $p_2<.001$) and confirmed us to use overall spelling ability as a proxy to control for influences of
25
26 general cognitive ability.
27
28
29
30
31
32

33
34 **Table A2:**

35
36
37 Cross-validated partial correlations (one-tailed) between bounded (upper panel) and
38
39 unbounded NLE task and basic arithmetic abilities
40
41

		A) Bounded NLE task				B) Unbounded NLE task			
Grade	Sample	Add	Sub	Mul	Div	Add	Sub	Mul	Div
5 th grade	Overall	-0.09	-0.20**	-0.08	-0.03	0.06	-0.14	0.03	-0.06
	Sub 1	-0.14	-0.31**	-0.11	-0.10	0.26	0.03	0.16	0.12
	Sub 2	-0.02	-0.16*	-0.06	0.04	0.18	0.12	0.33	0.26
6 th grade	Overall	-0.29**	-0.23**	-0.21**	-0.06	-0.08	-0.05	0.02	0.06
	Sub 1	-0.29**	-0.23**	-0.21*	-0.06	0.03	0.14	0.08	-0.08
	Sub 2	-0.35**	-0.29**	-0.29*	-0.12	-0.08	-0.01	-0.02	0.09
7 th grade	Overall	-0.43**	-0.44**	-0.24**	-0.25*	0.03	-0.17	0.14	0.05
	Sub 1	-0.52**	-0.39**	-0.26*	-0.33**	0.40	0.10	0.08	0.25
	Sub 2	-0.42**	-0.44**	-0.31*	-0.28**	0.19	-0.32	0.24	0.09

42
43
44
45
46
47
48
49
50
51
52
53
54
55 **Note:** Overall and cross-validated (Sub 1 and Sub 2) partial correlations were calculated for all grade
56
57 levels. Add=Addition, Sub=Subtraction, Mul=Multiplication, Div=Division. * $p<.05$, ** $p<.001$.
58
59
60
61
62
63
64
65

1 Second, partial correlations controlling for age and spelling skills were performed separately
2
3 for all grade levels. Table A2 in Appendix A2 shows the results of partial correlations between
4
5 number line estimation and basic arithmetic for the overall and the sub samples. Subsequently,
6
7 correlation coefficients in the two random sub-samples (n_1 and n_2) were compared to the
8
9 correlation coefficient of the overall sample (n) applying Student's t-distribution (Lenhard &
10
11 Lenhard, 2014). Comparisons revealed no significant differences in both bounded (for all
12
13 comparisons, $z > -.925$, $p > .10$) and unbounded number line estimation (for all comparisons,
14
15 $z < 1.635$, $p > .05$). This finding could prove the correlation coefficients in the overall sample as
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

REFERENCES

- Artemenko, C., Pixner, S., Moeller, K., & Nuerk, H. C. (2018). Longitudinal development of subtraction performance in elementary school. *British Journal of Developmental Psychology*, *36*(2), 188–205. <https://doi.org/10.1111/bjdp.12215>
- Ashcraft, M. H., & Moore, A. M. (2012). Journal of Experimental Child Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*(2), 246–267. <https://doi.org/10.1016/j.jecp.2011.08.005>
- 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>
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*(2), 545.
- 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 Presentations Influence Arithmetic Learning. *Child Development*, *79*(4), 1016–1031. <https://doi.org/10.1111/j.1467-8624.2008.01173.x>
- Bos, I. F. Den, Kroesbergen, E. H., Luit, J. E. H. Van, Xenidou-dervou, I., Jonkman, L. M., & Schoot, M. Van Der. (2015). Journal of Experimental Child Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology*, *134*, 12–29. <https://doi.org/10.1016/j.jecp.2015.02.002>
- Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin and Review*, *18*(12), 331–338. <https://doi.org/https://doi.org/10.3758/s13423-011-0059-z>
- Cohen, D. J., Blanc-goldhammer, D., & Quinlan, T. (2018). *A Mathematical Model of How People Solve Most Variants of the Number-Line Task*. 1–27. <https://doi.org/10.1111/cogs.12698>
- 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>
- Ebersbach, M., Luwel, K., & Verschaffel, L. (2015). The Relationship Between Children ' s Familiarity with Numbers and Their Performance in Bounded and Unbounded Number Line Estimations. *Mathematical Thinking and Learning*, *17*(2–3), 136–154. <https://doi.org/10.1080/10986065.2015.1016813>
- Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgements: The cyclical power model. *Psychological Review*, *107*(3), 500–524. <https://doi.org/10.1037/0033-295X.107.3.500>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.
- Huber, S., Moeller, K., & Nuerk, H.-C. (2012). Differentielle Entwicklung arithmetischer Fähigkeiten nach der Grundschule: Manche Schere öffnet und schließt sich wieder. *Lernen Und Lernstörungen*, *1*, 119–134.
- Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation.

1 *Developmental Psychology*, 53(6), 1088–1097. <https://doi.org/10.1037/dev0000305>

2
3 Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients.
4 *Communications for Statistical Applications and Methods*, 22(6), 665–674.
5 <https://doi.org/10.5351/CSAM.2015.22.6.665>

6
7 Lai, M., Zax, A., & Barth, H. (2018). Digit identity influences numerical estimation in children and adults.
8 *Developmental science*, 21(5), e12657.

9
10 Lenhard, W., & Lenhard, A. (2014). *Hypothesis tests for comparing correlations*.
11 <https://doi.org/10.13140/RG.2.1.2954.1367>

12
13 Link, T., Huber, S., Nuerk, H. C., & Moeller, K. (2014). Unbounding the mental number line-new evidence
14 on children's spatial representation of numbers. *Frontiers in Psychology*, 4, 1–12.
15 <https://doi.org/10.3389/fpsyg.2013.01021>

16
17 Link, T., Nuerk, H. C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic
18 competencies. *Quarterly Journal of Experimental Psychology*, 67(8), 1597–1613.
19 <https://doi.org/10.1080/17470218.2014.892517>

20
21 Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. (2009). Journal of Experimental Child Children ' s early
22 mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*,
23 103(4), 503–515. <https://doi.org/10.1016/j.jecp.2009.02.006>

24
25 Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth and Paladino.
26 *Developmental Science*, 14(5), 1194–1204. <https://doi.org/10.1111/j.1467-7687.2011.01070.x>

27
28 Opfer, J. E., Thompson, C. A., & Kim, D. (2016). Free versus anchored numerical estimation: A unified
29 approach. *Cognition*, 149, 11–17. <https://doi.org/10.1016/j.cognition.2015.11.015>

30
31 Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2016). Children's use of number line
32 estimation strategies. *European Journal of Psychology of Education*, 31(2), 117–134.
33 <https://doi.org/10.1007/s10212-015-0251-z>

34
35 Reinert, R. M., Hartmann, M., Huber, S., & Moeller, K. (2019). Unbounded number line estimation as a
36 measure of numerical estimation. *PLOS ONE*, 14(3), 1–16.
37 <https://doi.org/10.1371/journal.pone.0213102>

38
39 Reinert, R. M., Huber, S., Nuerk, H. C., & Moeller, K. (2017). Sex differences in number line estimation:
40 The role of numerical estimation. *British Journal of Psychology*, 108(2), 334–350.
41 <https://doi.org/https://doi.org/10.1111/bjop.12203>

42
43 Reinert, R. M., Huber, S., Nuerk, H., & Moeller, K. (2015). Strategies in unbounded number line
44 estimation? Evidence from eye-tracking. *Cognitive Processing*, 16(1), 359–363.
45 <https://doi.org/10.1007/s10339-015-0675-z>

46
47 Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental Number Line, Number Line Estimation, and
48 Mathematical Achievement: Their Interrelations in Grades 5 and 6. *Journal of Educational*
49 *Psychology*, 101(2), 359–372. <https://doi.org/10.1037/a0013840>

50
51 Schneider, M., Merz, J., Sticker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018).
52 Associations of number line estimation with mathematical competence: A meta-analysis. *Child*
53 *Development*. <https://doi.org/10.1111/cdev.13068> [Epub ahead of print]

54
55 Schneider, M., Thompson, C. A., & Rittle-Johnson, B. (2017). Associations of magnitude comparison and
56 number line estimation with mathematical competence: A comparative review. In *Cognitive*

57
58
59
60
61
62
63
64
65

1 *Development from a Strategy Perspective: A Festschrift for Robert Siegler* (pp. 100–119).
2 <https://doi.org/https://doi.org/10.4324/9781315200446>
3

4 Shapiro, S., & Wilk, M. B. (1965). *Biometrika Trust An Analysis of Variance Test for Normality (Complete*
5 *Samples) Published by: Oxford University Press An analysis of variance test for normality (complete*
6 *sa. 52(3), 591–611.*
7

8 Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child*
9 *Development, 75(2), 428–444.* <https://doi.org/10.1111/j.1467-8624.2004.00684.x>
10

11 Siegler, R. S., & Opfer, J. E. (2003). Research Article THE DEVELOPMENT OF NUMERICAL ESTIMATION:
12 Evidence for Multiple Representations of Numerical Quantity. *Psychological Science, 14(3), 237–*
13 *243.*
14

15 Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of
16 fractions. *Developmental Psychology, 49(10), 1994.*
17

18 Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between
19 number line estimation and mathematical achievement: The role of visuomotor integration and
20 visuospatial skills. *Journal of Experimental Child Psychology, 145, 22–33.*
21 <https://doi.org/10.1016/j.jecp.2015.12.004>
22

23 Slusser, E. B., & Barth, H. C. (2017). Intuitive proportion judgement in number-line estimation: Converging
24 evidence from multiple tasks. *Journal of Experimental Child Psychology, 162, 181–198.*
25

26 Slusser, E. B., Santiago, R. T., & Barth, H. (2012). Developmental Change in Numerical Estimation. *Journal*
27 *of Experimental Psychology: General, 142(1), 193–208.* <https://doi.org/10.1037/a0028560>
28

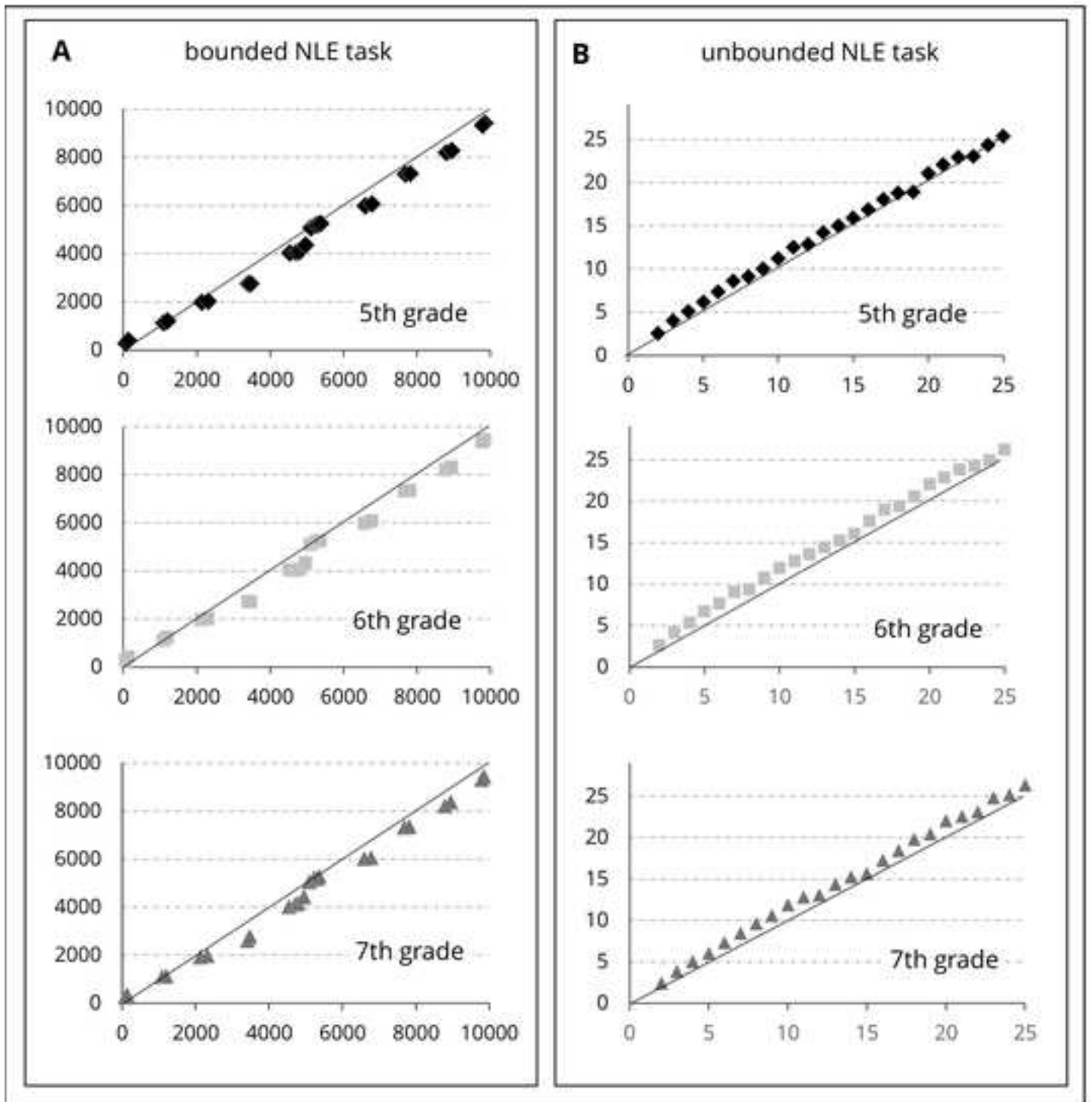
29 Thompson, C. A., & Opfer, J. E. (2010). How 15 Hundred Is Like 15 Cherries: Effect of Progressive
30 Alignment on Representational Changes in Numerical Cognition. *Child Development, 81(6), 1768–*
31 *1786.*
32

33 Young, L. K., & Booth, J. L. (2015). Student Magnitude Knowledge of Negative Numbers. *Journal of*
34 *Numerical Cognition, 1(1), 38–55.* <https://doi.org/10.5964/jnc.v1i1.7>
35

36 Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. D. Campbell
37 (Ed.), *Handbook of mathematical cognition* (pp. 331–345). New York: Psychology Press.
38

39 Zhu, M., Cai, D., & Leung, A. W. S. S. (2017). Number line estimation predicts mathematical skills:
40 Difference in grades 2 and 4. *Frontiers in Psychology, 8(SEP), 1–8.*
41 <https://doi.org/10.3389/fpsyg.2017.01576>
42

43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



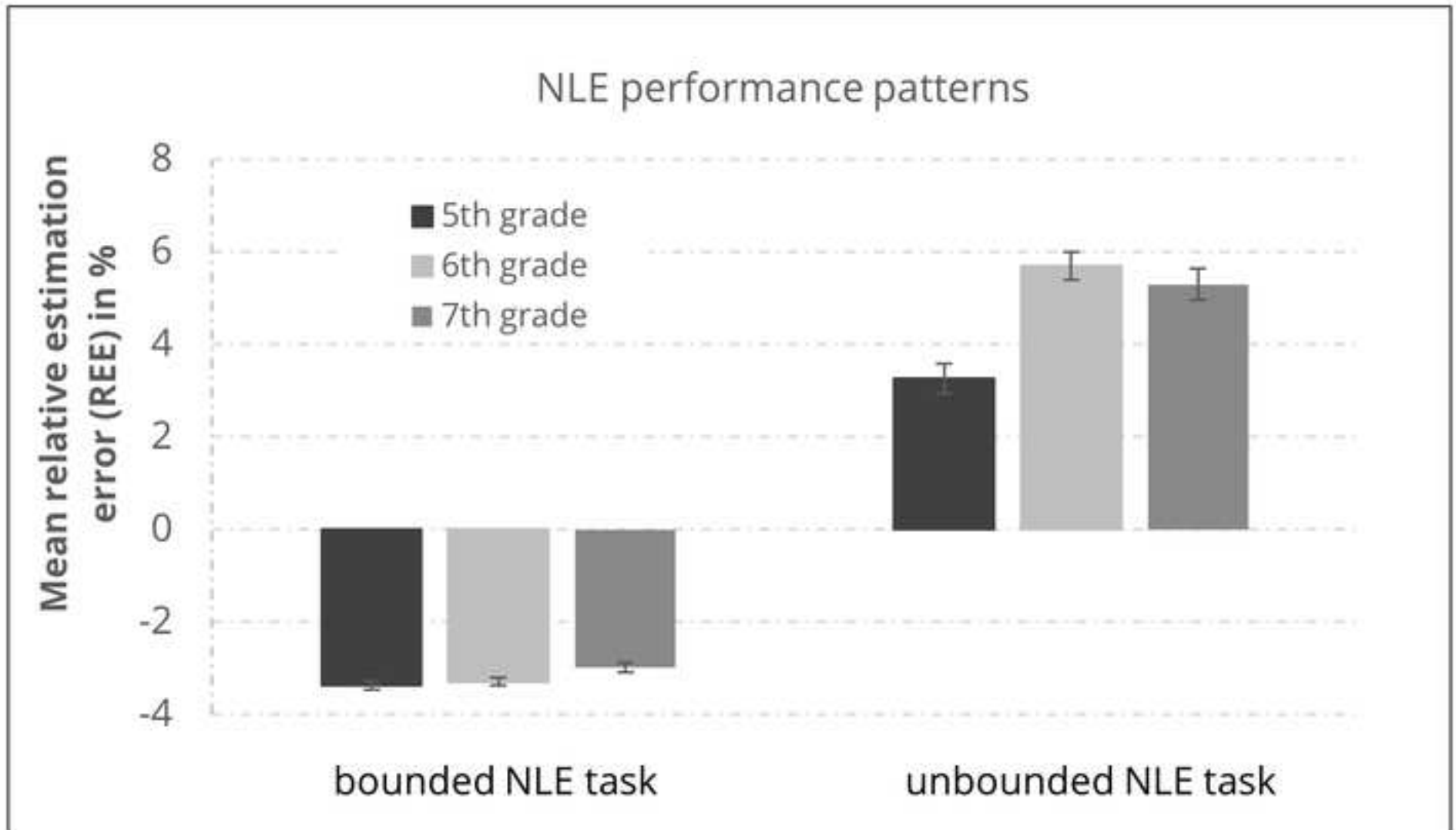


Figure 1

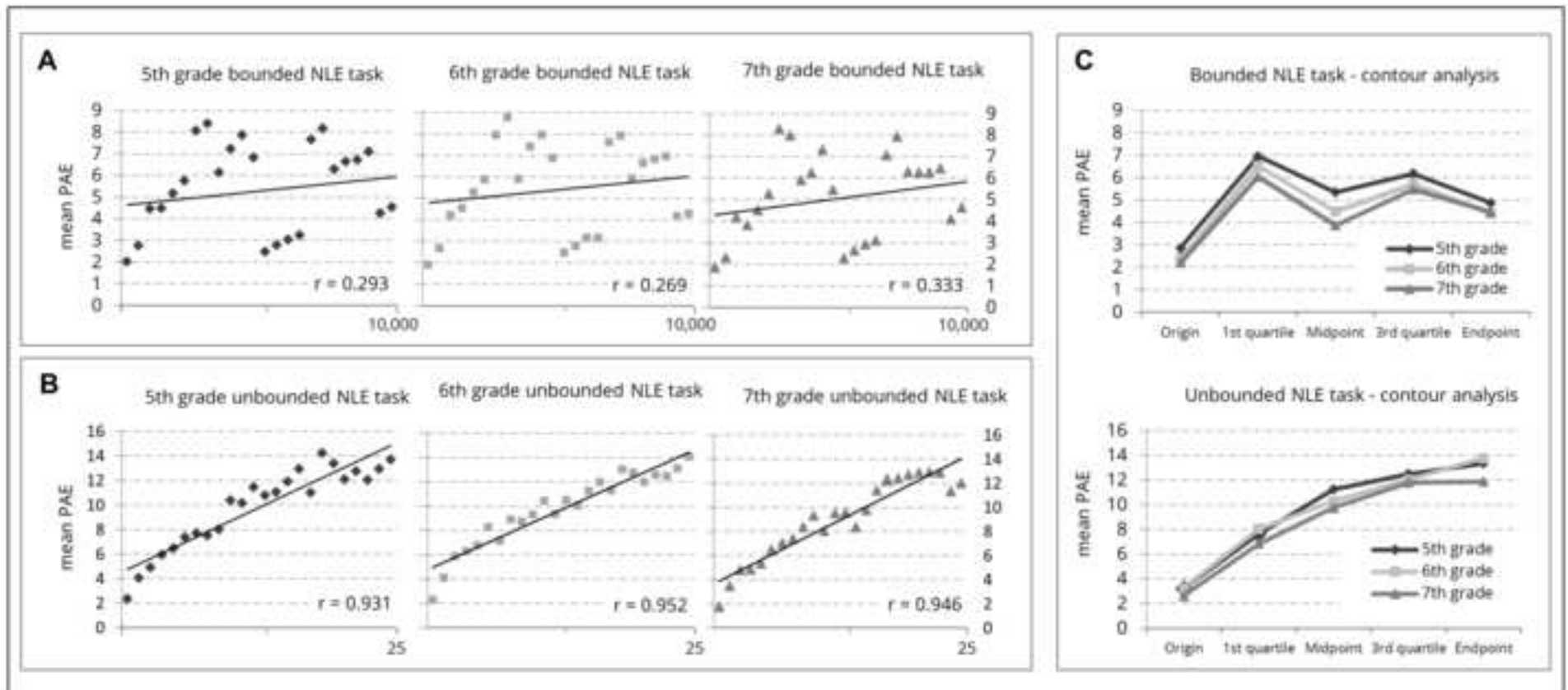


Figure 1: Mean estimation pattern for bounded (A) and unbounded (B) number line estimation: The figure shows target values (x-axis) and mean estimates (y-axis) for all three age levels, with $x=y$ in the diagonal.

Figure 2: Mean relative estimation errors (REE) for bounded and unbounded number line estimation performance.

Figure 3: Mean percentage absolute errors (PAE) for bounded and unbounded number line estimation performance. The left panel depicts mean PAE and target numbers in the (A) bounded and (B) unbounded NLE task separately for all grade levels. The right panel (C) shows contour analyses summarizing percent absolute errors (PAE) at specific reference points (cf. Ashcraft and Moore, 2012) for all grade levels (i.e., grade 5-7) again for bounded (top) and unbounded (bottom) NLE.