CHILD DEVELOPMENT

Child Development, September/October 2018, Volume 89, Number 5, Pages 1467–1484

Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis

Michael Schneider, Simon Merz, and Johannes Stricker
University of Trier

Bert De Smedt, Joke Torbeyns, Lieven Verschaffel, and Koen Luwel
University of Leuven

The number line estimation task is widely used to investigate mathematical learning and development. The present meta-analysis statistically synthesized the extensive evidence on the correlation between number line estimation and broader mathematical competence. Averaged over 263 effect sizes with 10,576 participants with sample mean ages from 4 to 14 years, this correlation was r = .443. The correlation increased with age, mainly because it was higher for fractions than for whole numbers. The correlation remained stable across a wide range of task variants and mathematical competence measures (i.e., counting, arithmetic, school achievement). These findings demonstrate that the task is a robust tool for diagnosing and predicting broader mathematical competence and should be further investigated in developmental and experimental training studies.

The number line estimation task is widely used in research on mathematical cognition, learning, and development. In the standard version of the task, on each trial, the participant is presented with an empty number line. Only the starting point and the endpoint are marked and labeled with the respective numbers. The participant is given a number, usually in the form of Arabic numerals, and is asked to locate the number on the line. Developmental studies use this task to trace the development of numerical magnitude understanding from early childhood over the elementary school years to adolescence. Educational studies evaluate interventions to improve performance on the number line estimation task as well as learning environments and curricula using the task to improve magnitude understanding and mathematical competence. This broad interest in the task might seem surprising, because number line estimation appears to be a simple and specific skill, which rarely plays a role in everyday life.

Among the reasons for the widespread use of the task is that studies found correlations between task performance and a wide range of other, more complex and advanced mathematical competence measures (see Siegler, 2016, for a review). For example, number line estimation has been found to correlate with counting (Östergren & Träff, 2013), arithmetic (Torbeyns, Schneider, Xin, & Siegler, 2015), and standardized school achievement tests (Ashcraft & Moore, 2012). In several studies, the correlation remained significant after controlling for potential confounding variables, such as parental income and education, race, ethnicity, working memory, intelligence, reading achievement, nonsymbolic numerical knowledge, proportional reasoning, and arithmetic proficiency (Bailey, Siegler, & Geary, 2014; Geary, 2011; Hansen et al., 2015; Hornung, Schiltz, Brunner, & Martin, 2014; Jordan et al., 2013; Östergren & Träff, 2013; Vukovic et al., 2014).

The correlation between number line estimation and broader mathematical competence is of theoretical as well as practical interest. Theories of numerical development can be evaluated by how well they are able to explain this hallmark finding. Educators may use the number line estimation task as a component of number sense tests or in mathematics curricula. For these purposes it would be useful to know how strong the correlation between number line estimation and mathematical competence actually is, how consistently it can be observed, and whether it is systematically higher for some task versions, subpopulations, mathematical competence measures, and so forth than for others. For example, theorists as well as practitioners could benefit from knowing whether the number line estimation

Correspondence concerning this article should be addressed to Michael Schneider, Educational Psychology, University of Trier - Division I, 54286 Trier, Germany. Electronic mail may be sent to m.schneider@uni-trier.de.

 0009-3920/2018/8905-0003
DOI: 10.1111/cdev.13068

task is of higher diagnostic value (i.e., more closely related with broader mathematical competence) for whole numbers than for fractions, for kindergartners than for middle school students, or for predicting counting rather than for predicting arithmetic. For these reasons, we conducted a meta-analysis on the correlation between number line estimation and broader mathematical competence. We combined all available effect sizes and examined the average effect size as well as moderating effects of third variables. In the next section, we describe the theoretical background before we introduce the moderator variables investigated in our meta-analysis.

Theoretical Background

A widely accepted theoretical explanation for the correlation between the number line estimation task and mathematical competence is that the number line estimation task assesses a central component of mathematical thinking, which aids the acquisition of broader and more advanced mathematical competence and thus correlates with measures of this competence. There are alternative accounts of what this central component might be.

According to one view in the literature, this central component is the representation of numerical magnitudes (Schneider, Grabner, & Paetsch, 2009; Schneider et al., 2008; Siegler & Opfer, 2003). This view is supported by an functional MRI study showing that the intraparietal sulcus, which is usually activated during numerical magnitude processing, is also activated during number line estimation (Vogel, Grabner, Schneider, Siegler, & Ansari, 2013). Further support comes from studies finding a developmental shift from logarithmic to linear estimate patterns, which is consistent with the view that the logarithmic estimate patterns reflect the logarithmic organization of number representations on the mental number line (Dehaene, Izard, Spelke, & Pica, 2008). The proficiency in representing and processing numerical magnitudes, which is assessed by number line estimation, might then support the acquisition of broader, more advanced mathematical competences. This beneficial influence of magnitude processing is a central assumption in several influential theories of mathematical learning and development, including Siegler's integrated theory of numerical development (Siegler, 2016; Siegler & Lortie-Forgues, 2014; Siegler, Thompson, & Schneider, 2011), Dehaene's (1997) account of number sense, and Spelke's (e.g., Feigenson, Dehaene, & Spelke, 2004) theory of mathematical core knowledge. Additionally, improvements in broader mathematical competence might improve numerical magnitude representation and number line estimation proficiency. This hypothesis is supported by a cross-lagged panel study, which found bidirectional predictive relations between number line estimation and a standardized mathematical achievement test (Friso-van den Bos et al., 2015).

According to another view in the literature, number line estimation mainly requires proportional reasoning, which relates the number to be estimated to the startpoint and the endpoint of the line. This notion is supported by studies finding that the estimates follow cyclical power functions (Barth & Paladino, 2011; Slusser, Santiago, & Barth, 2013), which are characteristic for proportion judgments (Hollands & Dyre, 2000). Number line estimation then correlates with broader mathematical competence, because proportional reasoning is a key component of competence in many mathematical domains (cf. Boyer, Levine, & Huttenlocher, 2008).

Alternatively or additionally to magnitude processing or proportional reasoning, the number line estimation might also be sensitive to spatial skills (Gunderson, Ramirez, Beilock, & Levine, 2012), visuomotor integration (Simms, Clayton, Cragg, Gilmore, & Johnson, 2016), measurement skills (D. J. Cohen & Sarnecka, 2014), counting strategies (Petitto, 1990), intelligence (Schneider et al., 2009), socioeconomic status (Ramani & Siegler, 2008), or other skills that support further mathematical learning and can thus explain the correlation between number line estimation and broader mathematical competence.

These accounts do not exclude each other. Several studies analyzing eye-tracking data, verbal strategy reports, or estimation patterns convergingly found that participants in a sample differ in their estimation patterns as well as in their estimation strategies (Peeters, Degrande, Ebersbach, Verschaffel, & Luwel, 2016; Petitto, 1990; Schneider et al., 2008; White & Szűcs, 2012). Testing how much the choice and execution of each strategy and the resulting estimate patterns are causally determined by magnitude representations, proportional reasoning, working memory, and other skills remains an important task for subsequent research. In the present meta-analysis, we could not tackle this task because there are too few studies on these questions, and these studies used heterogeneous methodological approaches. Instead, we focused here on the bivariate correlation between number line estimation and broader mathematical competence, because it has been investigated in many studies, each time in a similar way, so that the meta-analytically derived average correlations can be interpreted easily. Meta-analytic evidence on the average correlation and its moderators can then serve as starting point for further experimental and longitudinal studies on the underlying cognitive processes.

Magnitude Comparison as Benchmark

Like the number line estimation task, the magnitude comparison task is widely used in research on mathematical learning and development and is hypothesized to assess the mental representation and processing of numerical magnitudes (Ansari, 2008; De Smedt, Verschaffel, & Ghesquière, 2009; Dehaene, Dupoux, & Mehler, 1990). It thus provides a benchmark to compare findings obtained with the number line estimation task with. In magnitude comparison, the participants indicate which of two presented numerosities has the larger magnitude. In the most recent and largest meta-analysis, the correlation between magnitude comparison and mathematical competence was .24 averaged over 195 effect sizes obtained with nonsymbolic stimuli (i.e., dots) and .30 averaged over 89 effect sizes obtained with symbolic stimuli (i.e., Arabic numerals; Schneider et al., 2017). For nonsymbolic comparison, two smaller meta-analyses found similar correlations (Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014). Empirical studies (Hansen et al., 2017; Ye et al., 2016) and a recent qualitative review of the literature (Schneider, Thompson, & Rittle-Johnson, 2018) suggested that the correlation with mathematical competence might be stronger for number line estimation than for magnitude comparison, but this has not been tested meta-analytically so far. In the following section, we review variables that might moderate the correlation found with the number line estimation task in the current meta-analysis.

Possible Moderators of the Correlation With Mathematical Competence

Age

We expect that the correlation between number line estimation and broader mathematical competence increases with age, because the complexity of task demands and solution strategies increases with age. For example, young children typically estimate whole numbers in the 0–10 or 0–20 range and predominantly use counting-based strategies for estimating their locations. With increasing age, children can be presented with larger number ranges, can estimate fractions as well as whole numbers and will use more complex strategies, for example, proportional reasoning (e.g., locating 250 at 1/4 of the length of a 0–1,000 number line) or rounding a fraction to an easier to estimate number before trying to locate it on the line (Petitto, 1990; Siegler et al., 2011). Older children's more demanding tasks and strategies might more comprehensively assess the extent of their mathematical competence, thus leading to higher correlations. This hypothesis is not self-evident. Number line estimation seems to at least partly assess an understanding of numerical magnitudes. Learning about numerical magnitudes and their interrelations, for example, as in the counting sequence, is a central component of mathematical learning and competence tests during the preschool years but gets progressively less central in instruction and competence tests for older children (Siegler, 2016). Therefore, it is possible that the variance overlap between number line estimation and mathematical competence tests decreases with age. However, for the reasons outlined above, we still hypothesized to find increasing correlations with increasing age.

Number Type and Range

As explained above, the interpretation of any age differences needs to take into account that these are partly confounded with the types and ranges of the numbers to be estimated. The number types used in the published studies were whole numbers and fractions. Fraction estimation strategies require not only locating a magnitude on the line but also combining information from the numerator and denominator, and thus such strategies tend to be more complex than whole number estimation strategies (Rinne, Ye, & Jordan, 2017; Schneider & Siegler, 2010; Siegler et al., 2011). We hypothesized that this higher complexity of fractions might allow for a more fine-grained assessment of mathematical knowledge and skills, resulting in higher correlations with broader measures of mathematical competence for fractions than for whole numbers. In contrast, we did not predict systematic variations in the size of the correlation with respect to the range of the numbers to be estimated, because these are usually pragmatically chosen by researchers to avoid ceiling or floor effects in the age group under study. Therefore, averaged over studies, no systematic moderating effect of the number range was expected.

Variant of the Number Line Estimation Task

Several characteristics of the number line estimation tasks can easily be manipulated and result in tasks variants, which might differ in their correlations with mathematical competence. One such task characteristic is which positions on the number line are marked and labeled with the corresponding numbers. Typically, a bounded number line is used where the startpoint and the endpoint of the line (e.g., 0–100) are labeled. Less frequently, participants are presented with an unbounded number line without a labeled endpoint but with one unit given (e.g., the distance between 0 and 1; D. J. Cohen & Blanc-Goldhammer, 2011; Link, Nuerk, & Moeller, 2014). These studies are based on the assumption that bounded number lines elicit partly different cognitive processes than unbounded number lines. For example, the marked unit might invite counting strategies on unbounded number lines, and the labeled endpoint might invite proportional reasoning strategies in bounded number lines. Thus, the two task variants might differ in their correlations with mathematical competence. One can further distinguish between the number-to-position variant and the position-to-number variant of the task (Siegler & Opfer, 2003). In the former case, participants are presented with a number and have to locate its position on the line, whereas in the latter case, participants are given a position on a number line and have to estimate the corresponding number. Other and more peripheral task characteristics are the presentation medium of the task (i.e., paper-and-pencil vs. computerized), the physical length of the number line, the number of trials being presented to the participants, and the presentation mode of the number (i.e., printed digits, spoken number words, or dots). We had no hypotheses regarding these potential moderators. We still used them in explorative analyses, because from a practical point of view it would be helpful to know which task variants are most closely related to mathematical competence.

Index of Number Line Estimation Proficiency

A further variable to consider is the measure of proficiency on the number line estimation task. One measure is the percentage of correct trials, where an answer is considered as correct if it lies within a predefined interval (e.g., 10% of the line) around the correct position (e.g., Rittle-Johnson, Siegler, & Alibali, 2001). Another group of measures, the estimate deviation from the correct position, is based on the mean absolute difference between the correct position and the estimated position. This difference can be expressed in terms of percentage of the number line length (percentage of absolute error, PAE; e.g., Siegler & Booth, 2004) or in absolute terms (e.g., Geary, 2011). This measure is the most frequently used one, because it codes performance on each trial as a continuous score and thus yields more fine-grained results compared to the percentage of correctly solved trials, which is based on dichotomous coding of correct versus incorrect answers. A third index is obtained by plotting the estimated positions against the correct positions and computing the R^2 of a linear regression for these value pairs. Other and rarely used indices are the root mean square error (e.g., Anobile, Stievano, & Burr, 2013), which takes into account both estimate variance and bias, and composite measures that combine several of the previously described measures (e.g., Laski & Yu, 2014) to gain a more global assessment of number line estimation proficiency. As all indices of number line estimation proficiency are conceptually closely related, we expected the correlation between number line estimation and mathematical competence to be independent of the index used.

Mathematical Competence Measure

Most mathematical competence measures included in our meta-analysis differed in their content, which might lead to different associations with number line estimation performance. Mathematical competence was measured by: (a) counting tasks, (b) mental arithmetic tasks, (c) written arithmetic tasks, (d) school grades, and (e) standardized tests of mathematical achievement usually including several types of problems and aggregating their scores. To our knowledge, no previous study systematically compared how number line estimation relates to these measures. We therefore included this competence measure in our exploratory analyses to inform researchers and practitioners.

Temporal Order of the Assessments

A final difference between studies relates to the temporal order with which number line estimation performance and mathematical competence were measured. In cross-sectional designs both abilities are always measured at the same moment, whereas in longitudinal designs estimation performance can be assessed at T1 and mathematical competence at T2 (e.g., Jordan et al., 2013) or vice versa (e.g., Hornung et al., 2014). It remains an open question whether this temporal order affects the strength of the association between both abilities.

This Study

In sum, the number line estimation task is widely used in research on mathematical cognition, learning, and development, because it is assumed to assess a central foundation of mathematical thinking and correlates with many other mathematical tasks. However, there is a lack of knowledge on the exact strength of this relation, its consistency over studies, the breadth of conditions under which the relation can be found, and moderators that explain why the correlation was substantially higher in some studies than in others. We conducted a meta-analysis to investigate these points. The meta-analysis included six groups of moderator variables, which might affect the correlation as explained in the introduction section: (a) participant age, (b) number type and range, (c) task variant, (d) number line estimation index, (e) mathematical competence measure, and (f) temporal order of the assessments.

As previously outlined, we had five main hypotheses. First, the effect size for the association between number line estimation and mathematical competence was predicted to be significantly greater than zero when averaged over all available studies (Hypothesis 1). Second, the correlation was predicted to increase with age, as both task demands and the complexity of solution strategies tend to also increase with age (Hypothesis 2). Third, the correlation was predicted to be higher for fractions than for whole numbers, because fraction estimation is more demanding and complex than whole-number estimation (Siegler et al., 2011; Hypothesis 3). Fourth, the index of number line estimation proficiency was not expected to moderate the effect sizes, as the four types of measures are conceptually closely related to each other (Hypothesis 4). Finally, we hypothesized the correlation with mathematical competence to be stronger for number line estimation than for magnitude comparison, as suggested by Schneider et al. (2018; Hypothesis 5). In addition to testing these hypotheses, we performed a number of exploratory moderator analyses to investigate under which conditions number line estimation relates most closely to broader mathematical competence.

Method

Literature Search and Inclusion Criteria

We searched the title, abstract, and keywords of all articles in the literature database PsycINFO in February 2016 with the search string ((“math* achievement” or “math* competence” or “math* skill*” or “math* abilit*” or “math* performance” or “arithmetic*” or “num* skill*”) and (“number line*” or “numberline*” or “number-to-line” or “number-to-position” or “line-to-number” or “position-to-number”)) and limited the results to empirical studies with nondisordered human populations that had been published in a peer-reviewed journal in the English language. The search returned 141 hits. Unpublished results were not included, because they are hard to obtain. This might lead to nonrepresentative samples of unpublished studies (e.g., an overrepresentation of findings from the authors’ country or direct colleagues), which sometimes introduces new bias in a meta-analysis (Ferguson & Brannick, 2011). An additional explorative search returned 12 articles, so that we screened a total of 153 titles and abstracts for eligibility.

The inclusion criteria for our meta-analysis were: (a) The study reported original empirical findings (i.e., was not a re-analysis of already reported findings or a review). (b) The study included the number line estimation task either in the number-to-position or in the position-to-number version. The number line had to be empty except for a maximum of three labeled marks, because with more marks on the line it becomes less clear to what extent the participants estimated or simply read off the correct positions. In case of a *bounded number line* the startpoint and endpoint of the line were marked, and in case of an *unbounded number line* the startpoint and one unit on the line were marked. See Siegler and Thompson (2014) for an example of the rare case of three marks. (c) The study included a measure of mathematical competence other than number line estimation, that is counting, mental or written arithmetic, school grades, or a standardized test of mathematics achievement. Measures that are usually interpreted as assessing basic numerical processing (e.g., magnitude comparisons, same-different judgments, odd-even judgments, naming of magnitudes) were not considered as measures of mathematical competence because it is unclear to which extent they assess isolated and basic cognitive processes or a more general and directly school-relevant mathematical competence. (d) The study reported at least one standardized effect size of the strength and the direction of the bivariate relation between number line estimation proficiency and mathematical competence. The study also reported the sample size for this effect. Effect sizes from multivariate analyses (e.g., multifactorial analyses of variance or partial correlations) were excluded, because their outcomes depend on all variables included in the respective models, which limits the comparability. (e) The study reported at least one effect size for a sample with a majority of typically developing participants, who had not been diagnosed with dyscalculia or mathematical learning difficulties.

Two trained raters independently scanned the titles and abstracts of the found articles and decided for each one whether to exclude it or whether to obtain the full text for further inspection. A total of 74 full texts were obtained and then coded as either included or excluded. Interrater agreement for the inclusion of articles was 91%. Disagreements were resolved by discussion, leading to the inclusion of 41 studies in our meta-analysis.

Coding and Analyses

A trained coder extracted the information necessary for the meta-analysis from each included study. A second trained coder independently extracted 57 randomly chosen effect sizes with their moderator variables from the studies. Interrater agreement was 95% for the moderator variables and 100% for the effect size values. Again, disagreements were resolved by discussion. In the rare case that information vital for coding was missing or unclear in an article, we asked the authors to clarify by e-mail.

Prior to meta-analytic aggregation, all effect sizes were recoded so that a positive sign indicated that higher number line estimation proficiency was associated with higher mathematical competence. The effect sizes were corrected for measurement unreliability using Spearman's correction for attenuation (Hunter & Schmidt, 2004, p. 96) whenever the reliabilities of the measures were available and were left uncorrected otherwise. Two relatively high correlations reported by Östergren and Träff (2013) were obtained with measures with low reliabilities and would have had values larger than one after correcting for measurement unreliability. Because correlations greater than one are not defined, we entered these two correlations into our analyses without correcting them for unreliability. Following the advice by Hunter and Schmidt (2004, pp. 82/83), we did not subject the correlations to *Fisher Z* transformation before averaging them in our meta-analysis. Age group was coded as below 6 years of age (i.e., before the onset of formal school instruction on whole numbers in most countries), between 6 and 9 years (i.e., during whole-number instruction), or above 9 years (i.e., after whole-number instruction). Additionally, the sample mean age was coded as a continuous score.

As we included all relevant effect sizes from each study, the effect sizes were not statistically independent of each other. This would bias classical fixed-effects or random-effects meta-analyses. In particular, it would lead to an underestimation of the effect size variance in the population and, thus, to too narrow confidence intervals and too low error values for tests of the effect sizes against zero. We accounted for this problem by using a two-level regression model for the meta-analytic integration of the effect sizes. In this model, effect sizes on Level 1 were nested under independent samples on Level 2. The background and statistical details of multilevel regression models for meta-analyses are described by Hox (2002) and Van den Noortgate and Onghena (2003). We used inverse variance weighting so that effect sizes with smaller standard errors had greater weights in the meta-analysis. We entered most moderator variables as Level-1 predictors of effect sizes into our two-level model, because their values can differ between effect sizes within independent samples. Exceptions were moderators that mostly varied between studies and were thus entered as Level-2 predictors. The data were analyzed with the software MPlus 7.1 (Muthén & Muthén, 1998–2012). All reported confidence intervals are at the 95% level.

Results

Study Characteristics

The inclusion criteria were met by 41 articles (see Appendix S1). They reported results from 72 independent samples with 263 relevant effect sizes and 10,576 participants. All articles had been published in 2006 or after, indicating that research on the relation between number line estimation and mathematical competence is a young and quickly expanding field of research. After correction for measurement unreliability, the effect sizes ranged from −.196 to .860. The sample sizes were between 19 and 1,391 with a median of 99 (SD = 209).

The frequencies of the levels of the moderator variables are listed in Table 1. Of the 263 effect sizes, 19% had been found with participants younger than 6 years and, thus, before the onset of formal instruction on whole numbers in most countries; 42% had been found with participants aged 6–9 years; and 35% had been found with participants older than 9 years. Sample mean age ranged from 4 to 14 years (M = 8.39; SD = 2.66). About 67% of the 263 effect sizes were obtained using whole numbers and about 33% with fractions. The numerical ranges of the number lines were: 1 (in

Table 1

Correlation Between Number Line Estimation and Mathematical Competence for the Levels of the Moderator Variables

Moderator	r^+	[Lower 95% CI, upper 95% CI]	Samples	Effect sizes	Variance between samples	Variance within samples
Overall	.443	[.406, .480]	72	263	.016	.023
Age group (Level 1, R^2 = .145, p < .001)						
< 6 years	.296	[.253, .339]	11	50	.002	.023
6–9 years	.442	[.389, .495]	33	110	.009	.033
> 9 years	.491	[.434, .548]	27	91	.016	.013
Age, continuous (Level 1, R^2 = .073, p = .007)						
Number type (Level 2, R^2 = .144, p = .011)						
Whole numbers	.409	[.366, .452]	55	177	.013	.028
Fractions	.523	[.466, .580]	21	86	.012	.011
Numerical range (Level 1, R^2 = .001, p = .817)						
Number line type (R^2 = .112, p < .001)[a]						
Bounded	.447	[.410, .484]	72	253	.017	.017
Unbounded[a]	.055	[−.012, .122]	1	10	—	.011
Task type (R^2 = .001, p = .098)[a]						
Position to number[a]	.398	[.357, .439]	1	4	—	.002
Number to position	.444	[.407, .481]	71	259	.016	.023
Presentation medium (Level 2, R^2 = .013, p = .515)						
Computer	.460	[.411, .509]	37	124	.016	.010
Paper	.431	[.364, .498]	29	114	.019	.030
Physical line length (Level 2, R^2 = .029, p = .419)						
No. of number line estimation trials (Level 2, R^2 = .006, p = .615)						
Magnitude presentation in the number line task (Level 2, R^2 = .160, p = .004)						
Symbolic (digits)	.470	[.429, .511]	59	206	.015	.023
Nonsymbolic (dots)	.281	[.071, .491]	5	7	.045	.012
Spoken words	.398	[.357, .439]	—	4	—	.002
Several	.333	[.221, .445]	11	46	.006	.018
Index of number line estimation proficiency (Level 2, R^2 = .078, p = .147)						
% correct trials[a]	.451	[.351, .551]	2	6	—	.023
Estimate deviation	.450	[.407, .493]	50	166	.015	.021
Linear R^2	.441	[.365, .517]	22	49	.022	.012
Other	.351	[.233, .469]	12	42	.000	.046
Measure of math competence (R^2 = .054, p = .169)[a]						
Counting	.369	[.265, .473]	10	22	.013	.021
Mental arithmetic	.382	[.274, .490]	16	41	.010	.059
Written arithmetic	.466	[.405, .527]	25	62	.019	.004
Grades	.536	[.448, .624]	—	5	—	.016
Standardized tests	.468	[.413, .523]	39	108	.017	.020
Temporal order (Level 1, R^2 = .079, p < .001)						
Simultaneous	.427	[.384, .470]	62	191	.013	.028
Competence first	.538	[.477, .599]	17	39	.011	.010
Estimation first	.496	[.425, .567]	14	33	.011	.015

For categorical moderator variables with more than two levels, the lowest p value of the dummy-coded predictors is reported. Moderators for which no levels are listed were continuous.
[a]Estimated using a one-level regression model due to a too small number of sampling units.

10% of the effect sizes), 5 (6%), 10 (4%), 20 (8%), 30 (6%), 100 (41%), 1,000 (24%), 6,257 (1%), and 10,000 (1%). The ranges 1 and 5 were exclusively used with fractions. Because of the skewed distribution, numerical range was logarithmized before being used as predictor of effect sizes, leading to a min of 0, a max of 9.21, a mean of 4.29, and a *SD* of 2.18.

Most studies used the standard version of the number line estimation task (i.e., the number-to-position task with a bounded number line). Only 10 effect sizes were found with the unbounded number line and only 4 with the position-to-number version of the task. Forty-three percent of the studies used the paper-and-pencil version of the task, 47% used computers, and 10% did not report whether they used paper or computers. The physical line length in cm varied between 16.00 and 31.00 with a mean of 23.64 (SD = 3.18). The number of number line estimation trials in each study ranged from 6 to 44 (M = 21.80; SD = 8.51). Numbers in symbolic format were presented in 78% of the cases. Dots were presented in 3% of the cases (e.g., Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013), spoken number words in 2% (Ashcraft & Moore, 2012), and aggregated data from task versions with digits, dots, or spoken words in 18% (Muldoon, Towse, Simms, Perra, & Menzies, 2013).

Number line estimation proficiency was most frequently coded as relative or absolute difference between the estimated position and the correct position (63%), followed by proportion of variance of the estimates explained by a linear trend (19%), other measures (e.g., the standard deviation of the difference between the correct position and the estimated position; 16%), and percentage of correctly solved trials (i.e., the percentage of trials in with the position indicated by the participants lay in a predefined error interval around the correct position; 2%). The measures of mathematical competence were standardized mathematical achievement tests (45%), written arithmetic (26%), mental arithmetic (17%), and counting (8%). Only two studies (Schneider et al., 2009; Torbeyns et al., 2015) included school grades as competence measure. A longitudinal design was used in 27% of the cases.

Overall Effect for Number Line Estimation

The overall mean effect size and the mean effect sizes for the different levels of the categorical moderator analyses are listed in Table 1. The overall correlation between number line estimation and mathematical competence was r = .443 with a 95% confidence interval ranging from .406 to .480. The confidence interval did not include the zero, so that the effect size was statistically significant, which supports our Hypothesis 1. The variance of the effect sizes was 0.039 and their SD was 0.195. About 41% of the effect size variance was between samples, and 59% was within samples.

The 263 effect sizes did not deviate from a normal distribution, as indicated by a Kolmogorov–Smirnov test, p = .200. This symmetric distribution (see Figure 1) indicated the absence of a publication bias, which would have led to a right-skewed distribution (Egger, Smith, Schneider, & Minder, 1997). The absence of a publication bias in our database was also confirmed by Duval and Tweedie's (2000) trim-and-fill method. In this method, fictitious effect sizes are added to the left side of the effect size distribution until the ranks of the effect sizes distribute symmetrically, and a new overall effect size can be computed for the symmetric distribution. Our effect size distribution was already symmetric. So the trim-and-fill method left our results unchanged. Rosenthal's fail-safe N was 10,677. Only if this high number of unpublished studies with null results existed the number line-competence relation would

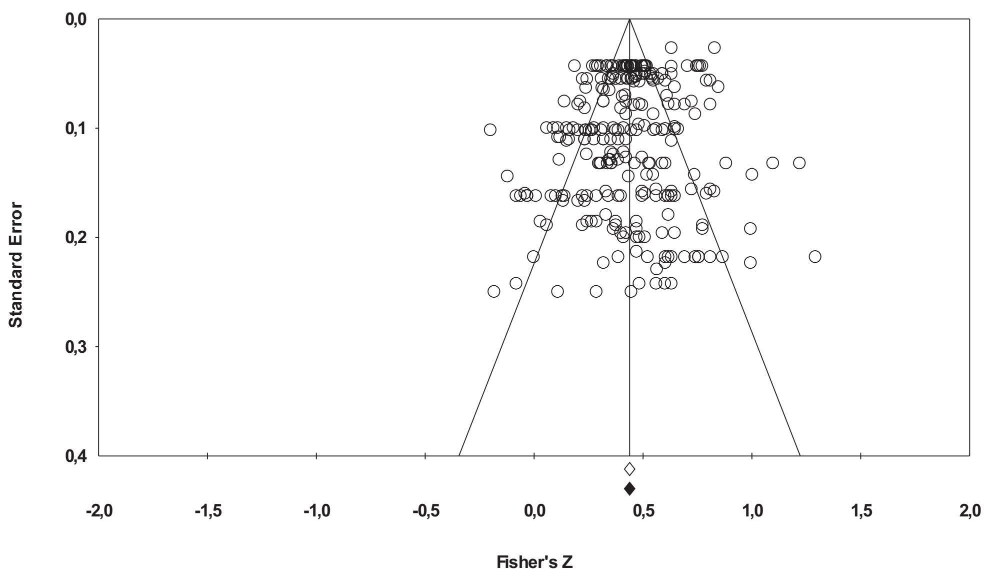

Figure 1. Funnel plot of the 263 effect sizes (here converted to Fisher's Z values) by standard error.

cease to be significant at the 5% level. Thus, the file-drawer problem is negligible in our case. The analyses also demonstrated that the results were not biased by an overly strong influence of specific samples. In a sensitivity analysis with the leave-one-out method, the omission of a sample never changed the overall correlation by more than $\Delta r = \pm .003$ points.

Age, Number Type, and Number Range as Moderators

The correlation between number line estimation and math competence was significantly moderated by the participants' age group. In line with Hypothesis 2, it was lowest for children younger than 6 years, higher for children aged 6–9 years, and highest for children older than 9 years. Age group as dummy-coded predictor of within-sample differences in effect sizes (i.e., as Level-1 predictor) explained 14.5% of the variance, $p < .001$, which is a medium strong effect by the commonly used standards of J. Cohen (1992). Two dummy variables were needed to code the information about three age group categories. So the regression returned two p values, one for each dummy variable. Here and in all similar analyses in the results section, we report the smallest of the p values along with the R^2 index of all predictors combined. Sample mean age in years as continuous predictor of effect size differences between studies explained a statistically significant variance proportion of 7.3%. The fact that age group explained about twice as much variance as continuous age indicates the nonlinearity of the moderation effect, which is visualized in Figure 2.

The correlation between number line estimation and mathematical competence was significantly moderated by the type of the numbers that had to be estimated. As predicted in Hypothesis 3, the correlation was higher for fractions than for whole numbers (see Table 1). This difference was significant with $p = .011$ and explained 14.4% of the variance of the effect sizes. As expected, number type and age group were not independent of each other, as only studies with children of 6 years and older used fraction estimation (see Table 2). For whole-number estimation, the correlation is highest for the 6- to 9-year olds and lower for younger and older children. For fraction estimation, the effect sizes were higher for children older than 9 years than for younger children. The numerical range of the line was unrelated to the effect sizes.

Task Variants and Measures as Moderators

The correlation was moderated by the variant of the number line estimation task. The correlation with mathematical competence was significantly positive for the standard, bounded form of the number line but not significantly different from zero for unbounded number lines, in which a unit on the line instead of the endpoint of the line is labeled with the corresponding number. This difference explained 11.2% of the variance of effect sizes in our meta-analysis. The correlation did not differ between the number-to-position version of the task and the position-to-number version. Presentation medium (paper vs. computer), physical line length, and number of estimation trials did not moderate the correlation. The correlation was highest when the numerical magnitude in the number line estimation task was presented as Arabic digits, lower for spoken number words, even lower for a mixture of several presentation formats (e.g., written digits and spoken words), and lowest for nonsymbolic magnitudes (i.e., dot patterns).

The index of number line estimation proficiency was not significantly related ($p = .147$) to the correlation between number line estimation and mathematical competence. This supports Hypothesis 4. The mean correlations were between .351 and .451 for all four types of measures. Descriptively, the correlations were highest and almost the same for estimate deviations from the correct position (e.g., PAE) and the percentage of correctly solved trials when using an error interval around the correct position. Linear R^2 and other measures were associated with descriptively slightly lower correlations.

The measure of math competence did not significantly moderate the estimation-competence relation, even though the correlations ranged from .536 for mathematics grades, over standardized mathematical achievement tests and arithmetic to .369 for counting.

The correlation was lower when the two variables were assessed at the same time and higher when they were assessed in longitudinal designs ($R^2 = .079$, $p < .001$). In the longitudinal studies, whether number line estimation was used as predictor of math competence over time or math competence was used as a predictor of number line estimation over time did not affect the effect sizes.

Magnitude Comparison

The number line estimation task and the magnitude comparison task are both widely used to index

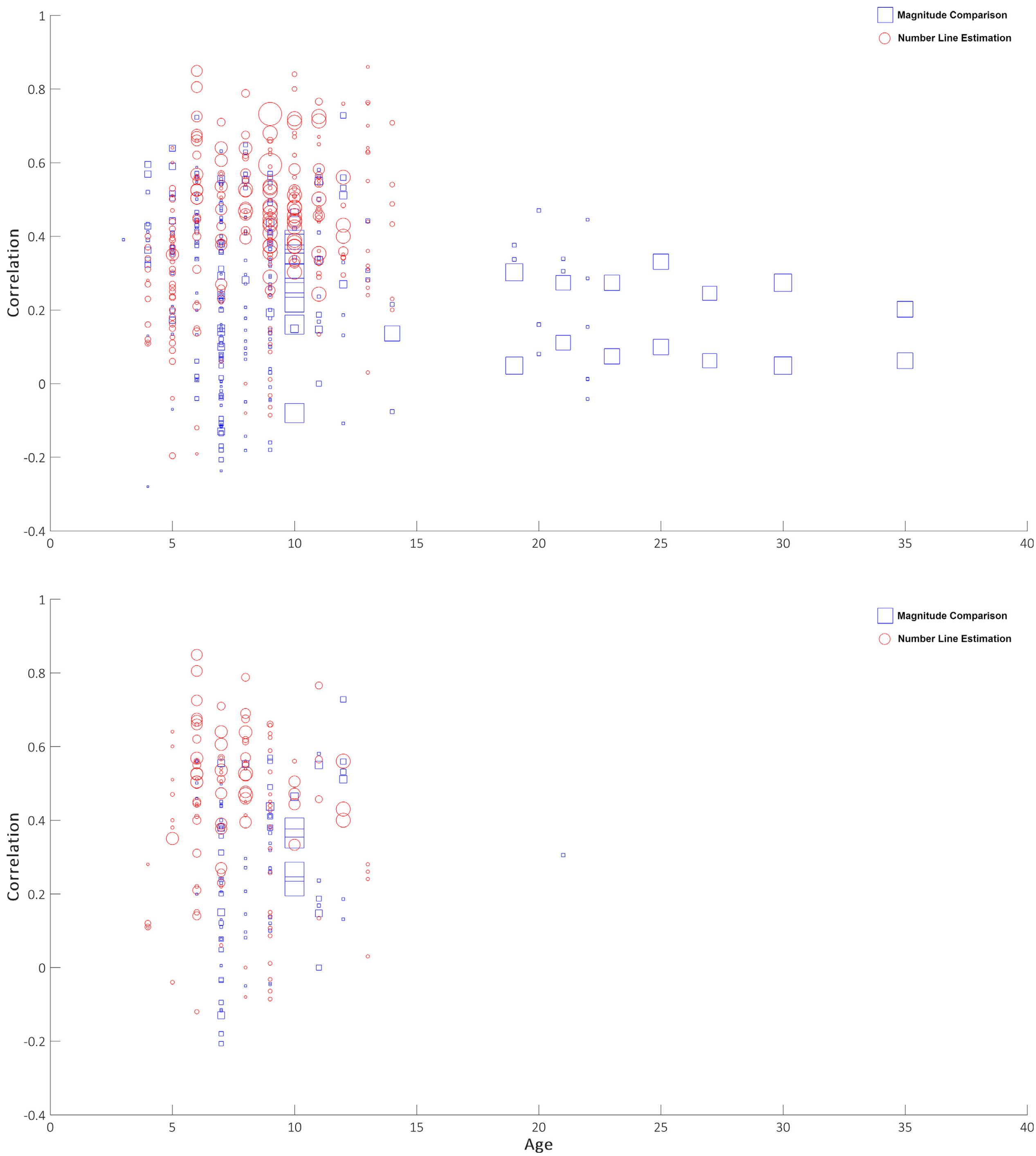

Figure 2. Correlations between magnitude processing and mathematical competence by age in years and task (estimation vs. comparison). The dot size is proportional to the sample size. Top: all effect sizes, bottom: effect sizes for symbolic whole numbers. [Color figure can be viewed at wileyonlinelibrary.com]

numerical magnitude processing and to predict mathematical competence. To be able to compare the correlations obtained with the two tasks, we merged the datasets from the present meta-analysis with the dataset from the most recent and largest meta-analysis on magnitude comparison and its correlation with mathematical competence (Schneider et al., 2017). This allowed us to directly compare the effect sizes in significance tests. The results are shown in Table 3.

Table 2
Correlations by Number Type and Age Group

	Whole numbers	Fractions
Younger 6 years		
r^+ and 95% CI	.296 [.253, .339]	—
Samples	11	0
Effect sizes	50	0
6–9 Years		
r^+ and 95% CI	.441 [.384, .498]	.454 [.409, .499]
Samples	32	2
Effect sizes	95	15
Older 9 years		
r^+ and 95% CI	.381 [.287, .475]	.529 [.470, .588]
Samples	9	21
Effect sizes	20	71

When all effect sizes included in the two meta-analyses were considered, the correlations were substantially higher for number line estimation than for magnitude comparison. In a meta-regression, the choice of task explained 29.1% of the variance in the 547 correlations. Thus, the past studies on number line estimation tended to find stronger associations with mathematical competence than the past studies on magnitude comparison. However, studies with the number line estimation task used fractions as well as whole numbers in a symbolic format, whereas studies with the magnitude comparison task mostly involved whole numbers in nonsymbolic and symbolic formats. When only the effect sizes obtained with whole numbers in symbolic format were considered, the difference between the two tasks became smaller but was still highly significant and explained 16.5% of the variance of these 266 effect sizes.

To examine possible interactions between type of task and participant age, we conducted the meta-regressions separately for the three age groups. Again, we included only the 266 effect sizes obtained with whole numbers in symbolic format. For children younger than 6 years, no comparison was possible due to a lack of effect sizes obtained with symbolic whole-number comparison. For the 6- to 9-year olds, the correlation with mathematical competence was substantially higher for number line estimation than for magnitude comparison. The choice of task explained an extremely high proportion of the effect size variance (R^2 = 36.7%) in this age group. In contrast, in persons older than 9 years, the choice of tasks was unrelated to the strength of the effect sizes (R^2 = 1.1%). Thus, the results support Hypothesis 5, that the correlation is higher for number line estimation than for magnitude comparison, only for the age group of 6- to 9-year olds.

Discussion

This study is the first meta-analysis on the association between number line estimation and broader mathematical competence. We found a substantial correlation between the two constructs, which was moderated by third variables. In the following, we first discuss the main findings with respect to our hypotheses, followed by possibly underlying mechanisms, and practical implications.

Main Findings

The meta-analytic results strongly support our Hypothesis 1, that number line estimation is associated with mathematical competence. Averaged over 263 effect sizes from 41 articles, the strength of the association was r = .443, which is a medium strong effect size by the standards of J. Cohen (1992). The 95% confidence interval from .406 to .480 indicated a good estimation precision. There was no evidence in favor of a publication bias. The correlation was also remarkably stable over the levels of the moderator variables. Table 1 lists these 29 levels (for number line estimation with whole numbers, with fractions, with bounded number lines, etc.). Sixteen of these 29 effect sizes are between .400 and .500. Twelve others are close to that interval and range from .281 to .538. Only one effect size, the one for unbounded number lines, is smaller and not statistically greater than zero. This consistency of the findings shows number line estimation to be a remarkably robust correlate and predictor of mathematical competence.

The meta-analytic findings also supported the three other hypotheses concerning moderating effects. The correlation increased with age (Hypothesis 2). This moderate increase was due to the more frequent use of fractions in older children (cf. Table 2). For whole numbers, the correlation was strongest during the elementary school years and slightly lower before and after. A possible explanation is that whole numbers, their magnitudes, and interrelations are central components of elementary school instruction, whereas earlier education has a stronger focus on the counting sequence and later education a stronger focus on algebra. For fractions, the effect size descriptively increased slightly from .454 for 6- to 9-year olds to .529 for older persons. Age-associated increases might partly be due to the

Table 3
Correlations r⁺ *of Number Line Estimation (From the Current Meta-analysis) and Magnitude Comparison (From Schneider et al., 2017) With Mathematical Competence*

	Number line estimation	Magnitude comparison	Difference	
			R^2	p
All numbers				
All age groups				
r^+ and 95% CI	.443 [.406, .480]	.274 [.239, .309]	.291	< .001
Samples	72	79		
Effect sizes	263	284		
Symbolic whole numbers				
All age groups				
r^+ and 95% CI	.408 [.365, .451]	.301 [.242, .360]	.165	.001
Samples	55	38		
Effect sizes	177	89		
Younger 6 years				
r^+ and 95% CI	.295 [.252, .338]	—	—	—
Samples	11	0		
Effect sizes	50	0		
6–9 Years				
r^+ and 95% CI	.438 [.379, .497]	.278 [.209, .347]	.367	< .001
Samples	32	24		
Effect sizes	95	66		
Older 9 years				
r^+ and 95% CI	.380 [.286, .474]	.345 [.249, .441]	.011	.646
Samples	9	14		
Effect sizes	20	23		

fact that larger number ranges are presented to older children. These larger ranges (e.g., number lines from 0 to 1,000) might tap a broader knowledge of numbers and allow for wider ranges of solution behavior than the simpler 0–10 number lines presented to younger children. The same might be true for fractions, where children usually start estimating simple unit fractions before progressing to more complex multidigit fractions. However, overall the number range did not moderate the effect sizes and the effect sizes did not monotonically increase with age (e.g., 6- to 9-year olds were better at whole number estimation than older children). This demonstrates that age-associated increases in the effect sizes cannot fully be attributed to age-associated increases in range of the presented numbers.

Fraction estimation was more closely related to mathematical competence than whole-number estimation, thus supporting Hypothesis 3. This finding can be explained by the greater complexity of fractions and fraction estimation strategies as compared to whole numbers and whole-number estimation strategies (cf. Siegler et al., 2011). The greater complexity might allow for wider ranges of solution behavior on the task (Rinne et al., 2017), which allows for a finer differentiation between children differing in their mathematical aptitude.

In line with Hypothesis 4, the correlation was not moderated by the number line estimation measure used. This demonstrates the conceptual similarity of the measures, all of which index in one way or another how close the estimated positions are to the correct positions on the line.

The explorative analyses showed that the correlation was higher for bounded than for unbounded lines. Among the possible explanations for this is that bounded number lines might be more familiar to children than unbounded number lines. Also, many participants use proportional reasoning strategies to position numbers on the line, but proportional reasoning is difficult or impossible on unbounded number lines (Link et al., 2014).

The correlation was moderated by the temporal order of the assessments. Unexpectedly, the correlation was lowest when both variables were measured at the same point in time and higher for the longitudinal studies. This contra-intuitive finding is hard to explain. Perhaps longitudinal studies used more reliable competence measures or more strict quality control (e.g., outlier cleaning) leading to higher correlations than less elaborate correlational one-shot studies. Within the longitudinal studies, the correlation was statistically significant for number line estimation as predictor of mathematical competence over time as well as for mathematical competence as predictor of number line estimation over time. This finding can be explained by assuming bidirectional causal relations between the two constructs (Friso-van den Bos et al., 2015). Further longitudinal studies carefully controlling for third variables and randomized controlled trials are needed.

The explorative moderator analyses also indicated that some task characteristics were unrelated to the correlation. These were the range of the presented numbers, whether the task was given on paper or on a computer screen, the physical length of the line, the number of estimation trials, and the measure of number line estimation proficiency. Notwithstanding the moderation effects discussed above, this demonstrates the general robustness of the task to small methodological variations. The correlation was also not moderated by the type of the mathematical competence measure, demonstrating that number line estimation is associated with a broad range of mathematical competence measures.

The similarity between the magnitude comparison task and the number line estimation task allowed us to use findings obtained with the former task as benchmarks for findings obtained with the latter task. As predicted in Hypothesis 5, the correlations found with the number line estimation task were higher than the correlations found with the magnitude comparison task. This was the case when all available effect sizes were considered, when all effect sizes obtained with symbolic whole number were considered, and when effect sizes obtained with symbolic whole numbers and 6- to 9-year-old children were considered. The finding was age-specific in that the advantage of number line estimation with symbolic whole numbers did not emerge in children older than 9 years.

Relations Between Estimation Patterns and Mathematical Competence

In the present meta-analysis, we focused on the correlation between overall number line estimation proficiency and mathematical competence, because this is what has been investigated in a large number of studies. Interestingly, two studies additionally reported associations between estimation patterns and mathematical competence. A study with 86 5- to 9-year olds and bounded as well as unbounded number lines fitted logarithmic and power functions to the estimation patterns. The coefficients of the logarithmic regression functions correlated descriptively stronger with addition and subtraction competence ($-.33 < r < -.42$) than the coefficients of the power functions ($-.10 < r < .27$; Kim & Opfer, 2017). This suggests that part of the correlation between number line estimation and math achievement can be explained by participants' use of linear versus logarithmic representations. However, a second study found in a sample of 124 elementary school students that the correlation between a standardized math achievement test and the proportion of explained variance was similar for linear regression functions ($.39 < r < .40$) and cyclical power functions ($.25 < r < .49$; Ashcraft & Moore, 2012). This indicates that proportional reasoning might also contribute to the correlation between estimation and achievement.

Underlying Causal Relations

These findings raise the question how the robust correlation between number line estimation and mathematical competence can be explained in terms of underlying causal relations. The present meta-analysis focused on correlational findings and did not allow the direct evaluation of hypotheses about causal relations. Any future investigation of these relations needs to consider two questions: First, which knowledge or skills are assessed by the number line estimation task and, second, how does this knowledge or these skills causally relate to broader mathematical competence?

With regard to the first question, there is unanimous evidence showing that participants do not somehow project numbers from their mental number line onto external number lines without any further processing. Error rates and estimation latencies (Ashcraft & Moore, 2012), estimate patterns (Siegler & Opfer, 2003), verbal reports (Peeters, Verschaffel, & Luwel, 2017), and eye tracking (Schneider et al., 2008; Sullivan, Juhasz, Slattery, & Barth, 2011) revealed that participants frequently use orientation points on the line. Whereas these might sometimes simply be recalled from memory (Sullivan & Barner, 2014), participants have also frequently been found to use rounding strategies and proportional reasoning strategies to find these orientation points. Additionally, participants use counting, addition, or subtraction strategies to estimate the position of a number relative to the startpoint, endpoint, or the nearest orientation point on the line (Link et al., 2014; Petitto, 1990; Schneider et al., 2008; Siegler et al., 2011). Thus, number line estimation proficiency reflects the proficiency in rounding, counting, proportional reasoning, and so on, at least to some extent.

The respective evidence is so strong that the question has been raised whether number line estimation might exclusively reflect these other mathematical skills and might be unrelated to numerical magnitude representation and processing (Barth & Paladino, 2011; LeFevre et al., 2013). However, rounding numbers, counting, proportional reasoning about numbers, and so on, require the processing and at least temporary mental representation of numerical magnitudes and thus depend on the quality of these processes and representations. Thus, the claim that number line estimation reflects proportional reasoning, landmark use, or any other strategy is compatible with the view that number line estimation assesses the processing and representation of numerical magnitudes, because proportional reasoning, landmark use, and other strategies operate on and thus require mental magnitude representations.

The involvement of that many component processes in number line estimation makes it hard to investigate the second open question, that is, what causal relations underlie the correlation between number line estimation and broader mathematical competence. Indirect evidence in favor of a causal effect of number line estimation proficiency on broader mathematical competence comes from longitudinal studies. Our meta-analytic results show that, averaged over 14 longitudinal studies, number line estimation was a statistically significant predictor of mathematical competence over time. Averaged over 17 longitudinal studies, mathematical competence was a statistically significant predictor of number line estimation over time. Several studies controlled these relations for possibly confounding variables and found that controlling weakened but did not eradicate the significant predictive relation (Bailey et al., 2014; Geary, 2011; Hornung et al., 2014; Jordan et al., 2013; Östergren & Träff, 2013; Vukovic et al., 2014). For example, number line estimation with whole numbers predicted fraction understanding in middle school in a sample of about 170 students after controlling for whole-number arithmetic proficiency, domain general cognitive abilities, parental income and education, race, and gender (Bailey et al., 2014). Because different studies controlled for different sets of variables, we could not meta-analytically synthesize these results.

Even more conclusive evidence on any causal relations would come from randomized controlled experiments in which the treatment group participates in a number line estimation training. The treatment group and the control group would need to complete a posttest, and ideally also a pretest, measuring broader mathematical competence, for example, in counting, arithmetic, or algebra. To our knowledge only one such experiment has been reported in the literature so far. This experiment included arithmetic as measure of mathematical competence, but did not find a statistically significant interaction effect between the test time (pretest vs. posttest) and the experimental groups (number line estimation, magnitude comparison, active control, passive control) on arithmetic (Maertens, De Smedt, Sasanguie, Elen, & Reynvoet, 2016). Another experiment found a causal effect of number line estimation training on children's memory for numbers but did not investigate whether this effect generalized to broader measures of mathematical competence (Thompson & Opfer, 2016).

Several other studies demonstrated the effectiveness of interventions, games, or curricula in which number lines were used in combination with other training elements, such as magnitude comparison, throwing a die or using a spinner and reading its number, adding numbers before estimating the position of the sum on a number line, or similar (e.g., Fuchs et al., 2013; Honoré & Noël, 2016; Thompson & Opfer, 2016). These studies found positive effects of the interventions on measures of mathematical competence. However, all studies left the question open whether these effects were caused by the number line or by other training components.

Several experimental training studies also showed that playing linear numerical board games can improve mathematical competence (Ramani & Siegler, 2008; Siegler & Ramani, 2009; Whyte & Bull, 2008). A similarity between these board games to number line estimation is that the participants have to map numbers (i.e., the number on the spinner) onto space (i.e., the number of fields they can move forward on the board). However, unlike in number line estimation, in numerical board games the players can simply count the fields they move forward, so that there is no estimation involved. In essence, there is indirect evidence for beneficial effects of number line trainings and related instructional interventions on broader mathematical competence. However, more direct evidence on the strength and direction of any causal relations between number line estimation and broader mathematical competence is needed.

Practical Implications

Notwithstanding the lack of direct evidence on causal relations, the present findings show that the number line estimation task is an easily applicable and robust tool for diagnosing and predicting broader mathematical competence. Individual differences in number line estimation proficiency correlate substantially with individual differences in counting, arithmetic, and standardized mathematical achievement tests. The correlation of $r = .443$ implies that 19.6% of the variance between persons in counting, arithmetic, and mathematical school achievement is associated with number line estimation proficiency. This association is stronger than the ones found with other important precursors of mathematical competence, including numerical magnitude comparison (Schneider et al., 2017) and working memory (Peng et al., 2016). Thus, in the absence of more detailed information, number line estimation performance can be used as a proxy for broader mathematical competences. At least three further characteristics of the number line estimation task contribute to its practical usefulness. First, the task takes little test time. Each trial of the number line estimation task requires only a few seconds to solve, and relatively small numbers of trials are necessary to obtain significant correlations with mathematical competence. The studies included here used, on average, only 21 trials. Second, the task allows for an assessment of mathematical competence, which is unbiased by differences in the participants' nonmathematical prior knowledge. It does not require any real-world knowledge, for example, about measurement units or physical objects (Booth & Siegler, 2006). Finally, the task is easy to administer and can flexibly be used in wide age ranges, on paper and on computer, in individual and group settings.

The number line estimation task correlates more strongly with mathematical competence than the magnitude comparison task does for all available effect sizes, for only effect sizes obtained with symbolic whole numbers, and, when holding age group constant, in the age group of 6- to 9-year olds. An explanation for the mostly higher correlation for number line estimation than for magnitude comparison could be that number line estimation assesses magnitude understanding on a continuous level, whereas magnitude comparison assesses magnitude understanding only on the ordinal level of larger/smaller judgments. The correlations between number line estimation and magnitude comparison were high in some studies (Laski & Siegler, 2007; Siegler et al., 2011), but low or heterogeneous in others (Sasanguie & Reynvoet, 2013; Schneider et al., 2009; Torbeyns et al., 2015), suggesting that it might sometimes be effective to use both number line estimation and magnitude comparison in competence tests and interventions, because the two tasks tap into partly different aspects of mathematical competence.

References

Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology*, *116*, 380–391. https://doi.org/10.1016/j.jecp.2013.06.006

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*, 278–291. https://doi.org/10.1038/nrn2334

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*, 246–267. https://doi.org/10.1016/j.jecp.2011.08.005

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*, 775–785. https://doi.org/10.1111/desc.12155

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125–135. https://doi.org/10.1111/j.1467-7687.2010.00962.x

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189–201. https://doi.org/10.1037/0012-1649.41.6.189

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developemental Psychology*, *44*, 1478–1490. https://doi.org/10.1037/a0013110

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, *18*, 331–338. https://doi.org/10.3758/s13423-011-0059-z

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, *50*, 1640–1652. https://doi.org/10.1037/a0035901

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 469–479. https://doi.org/10.1016/j.jecp.2009.01.010

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626–641. https://doi.org/10.1037/0096-1523.16.3.626

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*, 1217–1220. https://doi.org/10.1126/science.1156540

Duval, S., & Tweedie, R. (2000). Trim and Fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. https://doi.org/10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Ferguson, C. J., & Brannick, M. T. (2011). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, *17*, 120–128. https://doi.org/10.1037/a0024445

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. D. M. (2015). 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

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*, 683–700. https://doi.org/10.1037/a0032446

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539–1552. https://doi.org/10.1037/a0025510

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, *48*, 1229–1241. https://doi.org/10.1037/a0027433

Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development*, *35*, 34–49. https://doi.org/10.1016/j.cogdev.2015.02.001

Hansen, N., Rinne, L., Jordan, N. C., Ye, A., Resnick, I., & Rodrigues, J. (2017). Co-development of fraction magnitude knowledge and mathematics achievement from fourth through sixth grade. *Learning and Individual Differences*, *60*, 18–32. https://doi.org/10.1016/j.lindif.2017.10.005

Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, *107*, 500–524. https://doi.org/10.10371//0033-295X. 107.3.500

Honoré, N., & Noël, M.-P. (2016). Improving preschoolers' arithmetic through number magnitude training: The impact of non-symbolic and symbolic training. *PLoS ONE*, *11*, e0166685. https://doi.org/10.1371/journal.pone.0166685

Hornung, C., Schiltz, C., Brunner, M., & Martin, R. (2014). Predicting first-grade mathematics achievement: The contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Frontiers in Psychology*, *5*, 272. https://doi.org/10.3389/fpsyg.2014.00272

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781412985031

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*, *116*, 45–58. https://doi.org/10.1016/j.jecp.2013.02.001

Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. *Developemental Psychology*, *53*, 1088–1097. https://doi.org/10.1037/dev0000305

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*, 1723–1743. https://doi.org/10.1111/j.1467-8624.2007.01087.x

Laski, E. V., & Yu, Q. (2014). Number line estimation and mental addition: Examining the potential roles of language and education. *Journal of Experimental Child Psychology*, *117*, 29–44. https://doi.org/10.1016/j.jecp.2013.08.007

LeFevre, J.-A., Lira, C. J., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S.-L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, *4*, 641. https://doi.org/10.3389/fpsyg.2013.00641

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

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in preschoolers with comparison or number line estimation training: Does it matter? *Learning and Instruction*, *46*, 1–11. https://doi.org/10.1016/j.learninstruc.2016.08.004

Muldoon, K., Towse, J., Simms, V., Perra, O., & Menzies, V. (2013). A longitudinal analysis of estimation, counting skills, and mathematical ability across the first school year. *Developmental Psychology*, *49*, 250–257. https://doi.org/10.1037/a0028240

Muthén, B. O., & Muthén, L. K. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Östergren, R., & Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology*, *115*, 405–421. https://doi.org/10.1016/j.jecp.2013.03.007

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

Peeters, D., Verschaffel, L., & Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12233

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*, 455–473. https://doi.org/10.1037/edu0000079

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*, 55–78. https://doi.org/10.1207/s1532690xci0701_3

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394. https://doi.org/10.1111/j.1467-8624.2007.01131.x

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developemental Psychology*, *53*, 713–730. https://doi.org/10.1037/dev0000275

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*, 346–362. https://doi.org/10.1037/0022-0663.93.2.346

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*, 418–431. https://doi.org/10.1016/j.jecp.2012.10.012

Sasanguie, D., & Reynvoet, B. (2013). Number comparison and number line estimation rely on different mechanisms. *Psychologica Belgica*, *53*, 17–35. https://doi.org/10.5334/pb-53-4-17

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, *20*, e12372. https://doi.org/0.1111/desc.12372

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in grades 5 and 6. *Journal of Educational Psychology*, *101*, 359–372. https://doi.org/10.1037/a0013840

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, *23*, 424–437.

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1227–1238.

Schneider, M., Thompson, C. A., & Rittle-Johnson, B. (2018). Associations of magnitude comparison and number line estimation with mathematical competence: A comparative review. In P. Lemaire (Ed.), *Cognitive development from a strategy perspective: A festschrift for Robert Siegler* (pp. 100–119). London, UK: Routledge.

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, *19*, 341–361. https://doi.org/10.1111/desc.12395

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

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives*, *8*, 144–150. https://doi.org/10.1111/cdep.12077

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations. *Psychological Science*, *14*, 237–243. https://doi.org/10.1111/1467-9280.02438

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560. https://doi.org/10.1037/a0014239

Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful—Except when they're not. *Journal of Experimental Child Psychology*, *120*, 39–58. https://doi.org/10.1016/j.jecp.2013.11.014

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*, 273–296. https://doi.org/10.1016/j.cogpsych.2011.03.001

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

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*, 193–208. https://doi.org/10.1037/a0028560

Sullivan, J. L., & Barner, D. (2014). Inference and association in children's early numerical estimation. *Child Development*, *85*, 1740–1755. https://doi.org/10.1111/cdev.12211

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., & Barth, H. C. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review*, *18*, 557–563. https://doi.org/10.3758/s13423-011-0081-1

Thompson, C. A., & Opfer, J. E. (2016). Learning linear spatial-numeric associations improves accuracy of memory for numbers. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00024

Torbeyns, J., Schneider, M., Xin, Z., & Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. *Learning and Instruction*, *37*, 5–13. https://doi.org/10.1016/j.learninstruc.2014.03.002

Van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*, 765–790. https://doi.org/10.1177/0013164402251027

Vogel, S. E., Grabner, R. H., Schneider, M., Siegler, R. S., & Ansari, D. (2013). Overlapping and distinct brain regions involved in estimating the spatial position of numerical and non-numerical magnitudes: An fMRI study. *Neuropsychologia*, *51*, 979–989. https://doi.org/10.1016/j.neuropsychologia.2013.02.001

Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development*, *85*, 1461–1476. https://doi.org/10.1111/cdev.12218

White, S. L. J., & Szűcs, D. (2012). Representational change and strategy use in children's number line estimation during the first years of primary school. *Behavioral and Brain Functions*, *8*. https://doi.org/10.1186/1744-9081-8-1

Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology*, *44*, 588–596. https://doi.org/10.1037/0012-1649.44.2.588

Ye, A., Resnick, I., Hansen, N., Rodrigues, J., Rinne, L., & Jordan, N. C. (2016). Pathways to fraction learning: Numerical abilities mediate the relation between early cognitive competencies and later fraction knowledge. *Journal of Experimental Child Psychology*, *152*, 242–263. https://doi.org/10.1016/j.jecp.2016.08.001

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Articles Included in the Meta-analysis