





Conceptual Replication and Extension of the Relation Between the Number Line Estimation Task and Mathematical Competence Across Seven Studies

Alexa Ellis¹ , María Inés Susperreguy² , David J. Purpura¹ , Pamela E. Davis-Kean³ 

[1] Department of Human Development and Family Studies, Purdue University, West Lafayette, IN, USA. [2] Facultad de Educación, Pontificia Universidad Católica de Chile, Santiago, Chile. [3] Department of Psychology, University of Michigan, Ann Arbor, MI, USA.

Journal of Numerical Cognition, 2021, Vol. 7(3), 435–452, <https://doi.org/10.5964/jnc.7033>

Received: 2020-09-24 • Accepted: 2021-06-22 • Published (VoR): 2021-11-30

Handling Editors: Mojtaba Soltanlou, University of Surrey, Guildford, UK; Krzysztof Cipora, Loughborough University, Loughborough, UK

Corresponding Author: Alexa Ellis, Department of Human Development and Family Studies, Purdue University, 1202 West State Street, Room 336C, West Lafayette, IN 47907, USA. E-mail: alex@purdue.edu

Related: This article is part of the JNC Special Issue “Direct and Conceptual Replication in Numerical Cognition”, Guest Editors: Mojtaba Soltanlou & Krzysztof Cipora, Journal of Numerical Cognition, 7(3), <https://doi.org/10.5964/jnc.v7i3>

Supplementary Materials: Data, Materials [see Index of Supplementary Materials]



Abstract

A recent meta-analysis demonstrated the overall correlation between the number line estimation (NLE) task and children’s mathematical competence was $r = .44$ (positively recoded), and this relation increased with age. The goal of the current study was to conceptually replicate and extend these results by further synthesizing this correlation utilizing studies not present in the meta-analysis. Across seven studies, 954 participants, ranging from 3 to 11 years old (Age $M = 6.02$ years, $SD = 1.57$), the overall estimation-competence correlations were similar to those of the meta-analysis and ranged from $r = -.40$ to $-.35$. The current conceptual replication demonstrated that the meta-analysis captured a stable overall relation between performance on the NLE task and mathematical competence. However, the current study failed to replicate the same moderation of age group presented in the meta-analysis. Furthermore, the current study extended results by assessing the stability and predictive validity of the NLE task while controlling for covariates. Results suggested that the NLE task demonstrated poor stability and predictive validity in the seven samples present in this study. Thus, although concurrent relations replicated, the differential age moderation, lack of stability, and lack of predictive validity in these studies require a more nuanced approach to understanding the utility of the NLE task. Future research should focus on understanding the connection between children’s developmental progression and NLE measurement before further investigating the predictive and diagnostic importance of the task for broader mathematical competence.

Keywords

number line estimation, mathematical competence, numeracy, conceptual replication

The integrated theory of numerical development suggests that an understanding of numerical magnitudes underlies children’s mathematical achievement (Siegler, Thompson, & Schneider, 2011). Prior work has shown that there are a variety of reasons for why this may be, including neural overlap in tasks (Vogel, Grabner, Schneider, Siegler, & Ansari, 2013), developmental shifts in estimation patterns (Dehaene, Izard, Spelke, & Pica, 2008), proportional reasoning (Slusser, Santiago, & Barth, 2013), and the importance of space as a tool for understanding arithmetic concepts (Cipora, Patro, & Nuerk, 2015). Therefore, it has been widely accepted that our understanding of children’s numerical magnitude is important for numerical cognition research.



The Number Line Estimation Task

The number line estimation (NLE) task is a tool that has been widely used to assess children's numerical magnitude abilities and mathematical cognition (e.g., Fazio, Bailey, Thompson, & Siegler, 2014; Fuchs, Geary, Fuchs, Compton, & Hamlett, 2014; LeFevre et al., 2013; Lyons, Price, Vaessen, Blomert, & Ansari, 2014). Generally, this task presents children with an empty line where only the starting point and ending point are labeled (e.g., 0 on the far left of the line and 10 on the far right), and children are asked to mark where a given number goes on the number line. Multiple studies across a variety of disciplines have used this task as a measure of early mathematical understanding. Many of these studies vary, however, in the presentation of the task depending on characteristics of the children, such as child age (e.g., using nonsymbolic or symbolic as the number type, or using 1–10 or 0–100 as end points) (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Booth & Siegler, 2006; Reid, Baroody, & Purpura, 2015; Siegler & Booth, 2004; Thompson & Siegler, 2010). Regardless of presentation, moderate associations between the performance on the NLE task and a wide range of mathematical competence measures have been found across multiple studies (Siegler, 2016). Thus, this association is considered a robust finding and has important theoretical and practical implications in the study of mathematical cognition development and achievement.

Evidence for the association between performance on the NLE task and mathematical competence was strengthened by a recent meta-analysis (Schneider et al., 2018). Schneider and colleagues (2018) reverse coded the effect sizes present in the literature such that children who were more precise at placing numbers on the number line also showed higher scores on mathematical competence measures. The authors found that across 263 effect sizes, the correlation between the NLE task and broad mathematical competence was $r = .44$ [95% CI: 0.406, 0.480]. This relation increased with age such that the relation was $r = .30$ for children younger than 6 years of age, $r = .44$ for children 6–9 years of age, and $r = .50$ for children older than 9 years of age. Furthermore, the overall relation remained stable across task variants (e.g., number type, presentation medium, and number range) and mathematical measures, suggesting the NLE task is a robust correlate of mathematical competence.

The Present Study

The current study adds to the literature by leveraging data from seven independent and diverse studies not included in the previous meta-analysis (Schneider et al., 2018), to further examine and conceptually replicate the relation between the NLE task and mathematical competence. Specifically, this study aims to 1) replicate the effect sizes from the Schneider et al. (2018) meta-analysis, 2) replicate the age moderation, and lack of other moderations in the estimation-competence relation present in Schneider et al. (2018), and 3) extend their results by examining the stability and predictive validity of the NLE task while controlling for covariates.

Based on Schneider et al. (2018), our hypotheses are threefold: first, consistent with Schneider et al. (2018), we expect the effect size for the relation between the NLE task and mathematical competence to be significantly greater than zero, and to be moderate in size according to Cohen (1992). Second, also consistent with Schneider et al. (2018), we expect this relation to increase with age. Specifically, we expect weak to moderate ($r = .10$ to $.30$) relations between number line performance and math achievement for children under 6 years of age ($r = .30$; Schneider et al., 2018). However, in children above 6 years of age, we expect a moderate relation ($r = .30$ to $.50$) between number line performance and math achievement ($r = .44$; Schneider et al., 2018). Finally, number type, presentation medium, and number range are not expected to be significant moderators of the relation between NLE and mathematical competence. In addition to testing these replication hypotheses, we also performed exploratory analyses that extend the Schneider et al. (2018) study to investigate the stability and predictive validity of the NLE task while controlling for covariates. Thus, as these are exploratory, we do not have explicit hypotheses.

Method

Six studies for this conceptual replication were conducted in Midwestern cities in the United States of America, and one study was conducted in Chile. Studies were included in the analyses if they had administered a NLE task at least once. There were no data exclusion or outlier protocols. All study samples were collected between 2012 and 2020, whereas

Schneider et al. (2018) included studies between 2006 and 2018. Overall, the total sample (number of children that completed the NLE task at least at once) across all seven studies consisted of $N = 954$ children ranging from 3 to 11 years of age ($M_{age} = 6.02$ years, $SD = 1.57$).

For full descriptive information, see Table 1. Across all samples, 60% were younger than 6 years of age, 35% were between 6–9 years of age, and 5% were older than 9 years of age. Only 13% of the NLE tasks were nonsymbolic, whereas the rest (87%) were symbolic. Unlike Schneider et al. (2018), no number line tasks used in the current study utilized fractions as a number type. The number ranges used for the task were 10 (13%), 20 (60%), 100 (10%), or 1,000 (17%). All samples used the number to position approach with a bounded number line, and 62% of the sample completed the paper and pencil version, whereas 38% completed the task on a tablet.

Furthermore, all studies included at least one measure of mathematical competence beyond the NLE task. These measures included standardized mathematical achievement tests (Woodcock-Johnson Applied Problems subtest [WJAP; Woodcock, McGrew, & Mather, 2001], 69% of total sample; Preschool Early Numeracy Screener-Brief Version [PENS-B], 23% of total sample) or nonsymbolic number sense (Panamath; 45% of total sample). Forty-six percent of the sample were included in a longitudinal design and thus, also included a second measure of the NLE task.

Study 1

Participants

This study (Pathways) is an extension of a larger longitudinal project exploring the effects of schooling on executive functions. Data were collected on children attending four local elementary schools in Midwestern U.S. cities. The total sample includes three waves of children who were followed from kindergarten to second grade. Thus, there are three cohorts of children in the total sample. Data from this dataset were previously published (Ahmed, Grammer, & Morrison, 2021; Ellis et al., 2021), however, this publication is the only manuscript from this dataset to date that has examined research questions using the NLE task.

Schools included in this sample served children with a range of socioeconomic backgrounds based on school percentages of free and reduced-price lunch. Children were 4 to 6 years old ($N = 277$, $M_{age} = 5.71$ years, $SD = 0.40$, 53% male) and in kindergarten at the beginning of testing. Participants were primarily English speakers and had no known developmental disorders. Each child was assessed on a battery of executive function and achievement measures by a trained examiner. Around one year later, participants were assessed on the same measures ($n = 175$).

Number Line Estimation Task

All children completed a modified version of the NLE task. Children were presented with a piece of paper and pencil, and a line labeled 0 and 20 at the left and right ends, respectively. Their task was to draw a hash mark on the line indicating their best guess at where a given random integer between 0–20 fell on the line. The experimenter held one of three flipbooks that presented the child with the number they were to place on the line. Each flipbook contained the same randomly chosen 10 integers (1, 4, 6, 8, 9, 10, 13, 16, 17, and 19) in different orders. A new, unmarked line was used in each trial so that only one hash mark was placed on each line. Before testing, participants completed three practice trials on a shorter line labeled 0–10 to ensure an understanding of the instructions. The experimenter did not provide any feedback about children's placement of the hash marks during the practice trials. Participants then completed 10 test trials on the 0–20 line without feedback. To assess children's numerical magnitude accuracy, their percentage of absolute error (PAE) scores were calculated following the procedure described by Siegler and Booth (2004).

Mathematical Competence

Children completed the standardized English version of the Applied Problems subtest from the Woodcock-Johnson III Tests of Achievement (Woodcock et al., 2001). The Woodcock-Johnson III Tests of Achievement are standardized administrative tasks designed to provide information about a child's abilities compared to the national average. The Applied Problems subtest is a task in which children are presented with a set of questions to assess overall, broad mathematics abilities. Testing on this assessment was complete after six consecutive errors. This task is brief and can determine a wide range of mathematics abilities. It is also widely used in many nationally representative databases.

Table 1
Descriptive Statistics Across the Seven Studies

Variable	Study 1			Study 2			Study 3			Study 4			Study 5			Study 6			Study 7		
	N	M(SD)	Min Max	N	M(SD)	Min Max	N	M(SD)	Min Max	N	M(SD)	Min Max	N	M(SD)	Min Max	N	M(SD)	Min Max	N	M(SD)	Min Max
Age	275	5.71 (0.40)	4.93 6.94	92	5.55 (0.34)	4.92 6.67	35	5.65 (0.44)	4.83 6.42	62	6.92 (2.09)	3.42 11.50	101	4.31 (0.63)	3.17 5.25	261	7.88 (0.93)	5.75 10.41	124	4.18 (0.58)	3.12 5.26
Female	277	0.47 (0.50)	0 1	92	0.46 (0.50)	0 1	35	0.31 (0.47)	0 1	61	0.54 (0.50)	0 1	101	0.51 (0.50)	0 1	262	0.47 (0.50)	0 1	124	0.46 (0.50)	0 1
Mathematical Competence																					
WJAP Time 1	274	17.85 (4.17)	1 30	92	21.82 (4.11)	11 30	34	14.71 (4.25)	1 22	-	-	-	-	-	-	262	27.26 (5.09)	14 42	-	-	-
PENS-B Time 1	-	-	-	-	-	-	-	-	-	-	-	-	100	6.06 (4.90)	0 18	-	-	-	124	10.24 (5.82)	0 23
Panamath Time 1	250	80.84 (11.40)	33.33 98.86	-	-	-	-	-	-	56	84.56 (15.11)	25.00 100.00	-	-	-	-	-	-	-	-	-
WJAP Time 2	175	24.25 (3.76)	12 35	-	-	-	35	20.80 (4.75)	12 32	-	-	-	-	-	-	159	29.97 (4.59)	19 44	-	-	-
PENS-B Time 2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	114	13.55 (5.95)	0 24
Panamath Time 2	169	85.87 (7.81)	46.59 98.86	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Number Line Estimation Task																					
PAE Time 1	252	17.73 (10.18)	2.56 47.41	92	13.54 (7.63)	3.60 46.30	35	9.24 (4.81)	1.98 20.00	62	13.23 (10.06)	2.25 47.85	101	41.55 (11.98)	14.44 63.33	259	9.30 (9.13)	0.32 49.76	124	22.70 (7.09)	8.38 45.00
PAE Time 2	169	12.45 (7.11)	2.94 40.47	-	-	-	-	-	-	-	-	-	-	-	-	155	12.63 (7.60)	3.29 38.70	114	21.00 (8.90)	4.75 46.75
Number Type		Symbolic			Symbolic			Symbolic			Symbolic			Symbolic			Symbolic			Nonsymbolic	
Presentation Medium		Paper			Paper			Paper			Paper			Computer			Computer			Paper	
Number Range		0-20			0-20			0-20			0-20			0-10			0-100, 0-1,000			0-10	
Sample Name		Pathways			School Instruction			LENA			Sharing Task			Storybook			Chile			ANS	

Note. WJAP = Woodcock-Johnson III Applied Problems Subtest; PENS-B = Preschool Early Numeracy Screener-Brief; PAE = Percentage Absolute Error.

There are a total of 60 possible items in the Applied Problems subtest and an individual's score is the sum of their correct responses.

Children also completed a computerized nonsymbolic numerical discrimination task, titled Panamath (Halberda, Mazzocco, & Feigenson, 2008). During each trial, participants saw an array of yellow dots on one side of a computer screen and an array of blue dots on the other side; the dots were not displayed long enough to enable counting. Children indicated by button-press whether they thought there were more blue or yellow dots. No feedback was given on this task. The number of dots displayed varied, and the duration of the array presentation was adjusted for the participant's age. A blank screen appeared following the dot presentation, which persisted until the child pushed a button to indicate their response. Participants completed 16 trials (task duration was approximately one minute), and overall performance on the measure was calculated via the percentage of trials answered correctly.

Study 2

Participants

This sample (School Instruction) is part of a larger study exploring the role of classroom instruction on early math skills. Data from this study have not yet been previously published. Kindergarten children ($N = 92$, $M_{age} = 5.55$ years, $SD = 0.34$, 54% male) were recruited from four local elementary schools across 14 kindergarten classrooms in the greater southeast Michigan area. These four schools serve children with a range of socioeconomic backgrounds based on school percentages of free and reduced-price lunch. Participants received small gifts for their participation each time, parents and teachers received monetary compensation. Each child was assessed on a battery of math achievement measures by a trained examiner.

Number Line Estimation Task

Children completed the same version of the NLE task described in Study 1. Children's PAE scores were calculated based on the 10 test trials.

Mathematical Competence

Children completed the standardized mathematical achievement test (Woodcock-Johnson III Applied Problems subtest; Woodcock et al., 2001). The method for administering this test is described in Study 1.

Study 3

Participants

This sample (LENA) is part of a larger study exploring the role of the home environment on early numeracy skills. Data from the broader study were previously published at Susperreguy and Davis-Kean (2016); however, the NLE task was not included in the published article. These children were recruited from Midwestern U.S. cities. Children were 3–5 years old ($N = 35$, $M_{age} = 5.65$ years, $SD = 0.44$, 69% male) and in preschool at the time of testing. Each child was assessed on a battery of math achievement measures by a trained examiner.

Number Line Estimation Task

Children were presented with a number line on paper that ranged from zero to twenty. Children were required to select the appropriate position of a number on a number line between 0 and 20. First, children were shown where both 0 and 20 go in a horizontal line with "0" below the left end and "20" below the right end on a sheet of paper. Then, all numbers from 1 to 19 were presented, one at a time, in a random order, and they were asked to estimate the position of each number on the line, one number per number line. To assess the accuracy of the child's estimates, each child's PAE was calculated.

Mathematical Competence

Children completed the standardized mathematical achievement test (Woodcock-Johnson III Applied Problems subtest; Woodcock et al., 2001). The method for administering this test is described in Study 1.

Study 4

Participants

This sample (Sharing Task) was part of a larger project examining the relation between social division and math achievement in children. Data from this study have not yet been previously published. Children's ages ranged from 3 to 14 years old when they participated in the study ($N = 62$, $M_{age} = 6.92$ years, $SD = 2.09$, 46% male). Children were collected from a local museum and library in a Midwestern U.S. city. Each child was briefly assessed on a battery of math achievement measures by a trained examiner. Of interest in this study, children completed a version of the NLE task and Panamath (Halberda et al., 2008). Due to the testing environment in local libraries and museums where children were free to end the testing at any time, a portion of children only completed a portion of the tasks.

Number Line Estimation Task

Children completed the same version of the NLE task described in Study 1. Children's PAE scores were calculated based on the 10 test trials.

Mathematical Competence

Children completed the nonsymbolic numerical discrimination task (Panamath; Halberda et al., 2008). The method for administering this test is described in Study 1.

Study 5

Participants

This study (Storybook) recruited participants from a larger longitudinal project exploring a storybook intervention on children's mathematical skills. Data from this study have not yet been previously published. Participants included in the current study are the pretest data collected prior to the intervention random assignment. Children were recruited from local preschools in Midwestern U.S. cities. Children were 3 to 5 years old ($N = 101$, $M_{age} = 4.31$ years, $SD = 0.63$, 49% male) and in preschool at testing. All children had no known developmental disorders and were English speaking. Children were assessed on a battery of achievement measures by trained research assistants.

Number Line Estimation Task

Children completed an iPad version of the NLE task ranging from 0–10 (<https://hume.ca/ix/estimationonline.html>). The task presented children with a bounded number line and a number and they were asked to drag a hash mark to mark where they believed that number accurately belonged on the line. Children were asked to perform three practice trials (drag the line to 1, 4, and 9), and then they were presented with nine integers presented in a random order (1, 2, 3, 4, 5, 6, 7, 8, 9). Children's scores were calculated as the PAE across all non-practice trials.

Mathematical Competence

Children completed the PENS-B (Purpura, Reid, Eiland, & Baroody, 2015). The PENS-B is a 25-item early numeracy skill assessment that examines key mathematical domains identified in early preschool and kindergarten children. Test items assess children's counting skills, numerical relations, arithmetic operations, and numeral knowledge.

Study 6

Participants

This sample (Chile) is part of a more extensive cross-sequential study looking at the early predictors of math skills in Chilean children. Data from the broader study were previously published at [del Río et al. \(2020\)](#); however, the NLE task was not included in the published article. The children ($N = 263$, $M_{age} = 7.88$ years, $SD = 0.93$, 53% male, all Spanish speaking) were recruited across grade 1, grade 2, and grade 3 from five schools in Santiago, Chile, targeting both low and high socioeconomic status families. In the first wave of the study, children were assessed on a battery of early cognitive, linguistic, and numerical skills and math achievement measures by a trained examiner. Around one year later, participants were assessed on the same measures ($n = 159$).

Number Line Estimation Task

An iPad version of the NLE task was administered, similar to Study 5 (<https://hume.ca/ix/estimationline.html>; ranging from 0–100 in grade 1, and 0–1,000 in grades 2 and 3). Children were presented with a bounded number line and a number and had to drag a hashmark to mark where they believed that the number accurately belonged on the line. Children completing the number line ranging from 0–1,000 saw 24 integers presented in a random order (6, 18, 59, 97, 124, 165, 211, 239, 344, 383, 420, 458, 542, 580, 617, 656, 761, 789, 835, 876, 903, 941, 982, 994). Those who completed with the 0–100 number line saw 22 integers presented in a random order (2, 3, 5, 8, 12, 17, 21, 26, 34, 39, 42, 46, 54, 58, 61, 67, 73, 78, 82, 89, 92, and 97). Children's score was calculated as the PAE across all possible trials.

Mathematical Competence

Children completed the Spanish version of the Woodcock-Johnson III Applied Problems subtest (Batería III Woodcock-Muñoz; [Muñoz-Sandoval, Woodcock, McGrew, & Mather, 2005](#)). This task was administered similarly to that described in Study 1.

Study 7

Participants

This sample (ANS) is part of a larger study on children's mathematical, executive function (EF), and literacy development during preschool. Data from the broader study were previously published at [Purpura and Logan \(2015\)](#), and [Purpura and Simms \(2018\)](#); however, the NLE task was not included in these published articles. Preschool children ($N = 124$, $M_{age} = 4.18$ years, $SD = 0.58$, 54% male) were recruited from 12 different preschools in Midwestern U.S. cities. Families from a broad range of socioeconomic status were recruited (36% < 4-year college degree, 21.6% 4-year college degree, 42.4% > 4-year college degree). Each child was assessed on a battery of math achievement measures by a trained examiner in the fall of preschool. During the spring of the same year, participants were assessed on the same measures ($n = 114$).

Number Line Estimation Task

Children completed a modified version of the NLE task with paper and pencil designed to be a non-verbal number line task where, rather than being presented with numbers, children were presented with sets of dots (1–10). The task also included modifications according to [Reid et al. \(2015\)](#). For example, the set to be represented was presented on a flashcard instead of above the middle of the number line, the number line had a marker at 0, 1, and 10, and used the example of a rabbit hopping was provided to ensure that children understood the task. Therefore, this number line task included three benchmarks as a bounded condition.

Mathematical Competence

Children completed the nonsymbolic numerical discrimination task (Panamath) and the standardized numeracy skill assessment (PENS-B). The methods for administering these tests are described in Study 4 and Study 5, respectively.

Analytic Approach

Data analyses were run using the *psych* package (version 2.0.7) in *RStudio* (version 1.1.456). Syntax and a simulated dataset is available at <https://osf.io/qswav/>. Using the seven different studies, we examined the overall correlation between children's performance on a given version of the NLE task and their mathematical achievement.

Further, we tested if different aspects of the NLE and mathematical competence measures moderated any correlations using a *z* test (Soper, 2020). These possible moderators included age group (below 6, 6–9, above 9), number range (0–10, 0–20, 0–100, or 0–1,000), presentation medium (computer or paper), and number type (symbolic or nonsymbolic). Although Schneider et al. (2018) assessed multiple moderators in the relation between performance on the NLE task and mathematical competence, the samples in this manuscript did not include all of the moderators included in their analyses. For example, Schneider et al. (2018) used number line type (bounded or unbounded), task type (position to number, or number to position), and index of NLE proficiency (PAE, estimate deviation, or linear R^2). However, all samples in this manuscript used bounded number lines, position to number task type, and PAE as an index of NLE proficiency. Thus, these moderators were excluded from analyses.

Finally, exploratory multiple linear regressions were conducted to extend the results from Schneider et al. (2018) by examining the stability and predictive validity of the NLE task. These regressions tested whether the NLE task was stable across time, stable while controlling for other mathematical competence measures, and whether the NLE task predicted other mathematical measures while controlling for initial mathematical scores. Schneider et al. (2018) examined the predictive validity of the NLE task for later competence, however, the authors were unable to control for other key variables (e.g., participant age) given the meta-analytic data. Therefore, we extend these results by focusing on the stability and predictive validity of the task while controlling for covariates.

Sensitivity Power Analyses

The seven studies used in this manuscript are existing datasets, thus, a sensitivity power analysis was used to calculate the range of minimally detectable effect sizes (MDES) given the sample sizes across the proposed correlations (Cribbie, Beribisky, & Alter, 2019; Giner-Sorolla et al., 2019). Across all proposed moderators, the highest sample size was $n = 634$ and the smallest sample size was $n = 71$. G*Power was used to run sensitivity power analyses given this range of sample sizes necessary to detect true significant effect sizes (Faul, Erdfelder, Lang, & Buchner, 2007). A bivariate correlation with 634 participants, $\alpha = 0.05$, and power $(1-\beta) = 0.80$, the sensitivity power analysis suggested that the MDES was 0.11 (Faul et al., 2007). For the smallest sample size, a bivariate correlation with 71 participants, $\alpha = 0.05$, and power $(1-\beta) = 0.80$, the sensitivity power analysis suggested that the MDES was 0.29 (Faul et al., 2007). Thus, our sensitivity power analyses suggest that our larger dataset is powered to detect significant effect sizes as low as 0.11 for our highest sample and as low as 0.29 for our lowest sample.

Results

Overall Estimation-Competence Relation

The overall effect sizes and effect sizes by moderator variables are listed in Table 2. Although Schneider et al. (2018) recoded all effect sizes such that a positive sign indicated higher scores on the NLE task was associated with higher mathematical competence, we chose to report our effect sizes true to the literature such that lower scores on the NLE task (PAE) were associated with higher mathematical competence. The overall correlation between the NLE task and mathematical competence ranged from $r = -.40$ to $-.35$ (95% CIs ranged from -0.51 to -0.26). The 95% confidence intervals did not include zero across all mathematical competence measures, suggesting the relation was statistically significant. Thus, these results support the first replication hypothesis that the NLE task is significantly associated with mathematical competence.

Table 2*Correlations Between Number Line Estimation and Mathematical Competence Measures*

Moderator	<i>r</i>	95% CI	<i>n</i>	Studies
Overall				7
WJAP	-.36	[-0.42, -0.30]	634	4
PENS-B	-.40	[-0.51, -0.30]	224	2
Panamath	-.35	[-0.44, -0.26]	408	3
PAE Time 2	.24	[0.13, 0.33]	409	3
Age group				
<6 years				7
WJAP	-.40	[-0.50, -0.29]	306	4
PENS-B	-.40	[-0.51, -0.28]	224	2
Panamath	-.29	[-0.39, -0.29]	320	3
PAE Time 2	.23	[0.11, 0.33]	226	3
6–9 years				5
WJAP	-.11	[-0.19, -0.01]	291	4
PENS-B	–	–	–	–
Panamath	-.20	[-0.42, 0.04]	75	2
PAE Time 2	-.05	[-0.23, 0.14]	180	2
Number Type				
Symbolic				6
WJAP	-.36	[-0.41, -0.30]	634	4
PENS-B	-.26	[-0.45, -0.06]	100	1
Panamath	-.27	[-0.40, -0.14]	285	2
PAE Time 2	.08	[-0.05, 0.22]	295	2
Non-Symbolic				1
WJAP	–	–	–	–
PENS-B	-.22	[-0.38, -0.05]	124	1
Panamath	-.26	[-0.39, -0.10]	123	1
PAE Time 2	.06	[-0.15, 0.23]	114	1
Presentation Medium				
Computer				2
WJAP	.06	[-0.04, 0.18]	259	1
PENS-B	-.26	[-0.45, -0.08]	100	1
Panamath	–	–	–	–
PAE Time 2	-.01	[-0.19, 0.16]	151	1
Paper				5
WJAP	-.4	[-0.48, -0.30]	258	3
PENS-B	-.22	[-0.38, -0.07]	124	1
Panamath	-.35	[-0.44, -0.25]	408	3
PAE Time 2	.23	[0.11, 0.34]	258	2
Number Range				
0 to 10				2
WJAP	–	–	–	–
PENS-B	-.40	[-0.50, -0.27]	224	2
Panamath	-.26	[-0.40, -0.10]	123	1
PAE Time 2	.06	[-0.13, 0.25]	114	1

Moderator	<i>r</i>	95% CI	<i>n</i>	Studies
0 to 20				4
WJAP	-.40	[-0.49, -0.31]	375	3
PENS-B	-	-	-	-
Panamath	-.27	[-0.39, -0.14]	285	2
PAE Time 2	.19	[-0.02, 0.38]	144	1
0 to 100				1
WJAP	-.51	[-0.63, -0.39]	95	1
PENS-B	-	-	-	-
Panamath	-	-	-	-
PAE Time 2	.53	[0.36, 0.68]	80	1
0 to 1,000				1
WJAP	-.55	[-0.64, -0.46]	164	1
PENS-B	-	-	-	-
Panamath	-	-	-	-
PAE Time 2	.57	[0.35, 0.73]	71	1

Note. WJAP = Woodcock-Johnson III Applied Problems Subtest; PENS-B = Preschool Early Numeracy Screener-Brief; PAE = Percentage Absolute Error.

Moderators

Measure of Mathematical Competence

The tests of moderation for the measures of mathematical competence were not found to be statistically significant for the NLE relation (see Table 3 for all *z* statistics). The overall relation between NLE and mathematical competence measure ranged from $r = -.40$ to $-.35$.

Age

Due to our small sample size for children in the above 9 age group ($n = 35$), we were unable to test conceptual replication results for this group from the Schneider et al. (2018) meta-analysis. However, we were able to test the other two age moderating groups presented in the original meta-analysis. Children below 6 years of age demonstrated a moderate relation between both standardized mathematical measures and nonsymbolic number sense and performance on the NLE task, $r(\text{WJAP}) = -.400$; $r(\text{PENS-B}) = -.403$; $r(\text{Panamath}) = -.292$. Relations then decreased for children between the ages of 6–9 years, $r(\text{WJAP}) = -.106$; $r(\text{PENS-B}) = NA$; $r(\text{Panamath}) = -.201$.

The test of moderation for the correlation between the NLE task and mathematical competence was found to be statistically significant for the participants' age group on the Applied Problems subtest measure ($z[595] = -3.86$, $p < .001$; see Table 3). However, this moderation by age group was not as hypothesized based on the Schneider et al. (2018) meta-analysis. The correlations were highest among children younger than 6 years of age, and lower for children aged 6–9 years. Interestingly, the test of moderation for the estimation-competence relation was not found to be statistically significant by age group for the other competence measure (Panamath; $z[393] = -0.74$, $p = .46$). In sum, the estimation-competence relation demonstrated an age moderation for the Applied Problems subtest measure, but not the Panamath measure.

Number Type

The tests of moderation for the type of numbers presented to children were also not found to be statistically significant for the correlation between performance on the NLE task and mathematical competence (see Table 3). In line with our third hypothesis, relations among mathematical competence and performance on the NLE task were similar across both symbolic, $r(\text{WJAP}) = -.361$; $r(\text{PENS-B}) = -.260$; $r(\text{Panamath}) = -.267$, and nonsymbolic, $r(\text{WJAP}) = NA$; $r(\text{PENS-B}) = -.222$; $r(\text{Panamath}) = -.262$, number types estimated.

Table 3*Correlation Comparisons by Moderator and Mathematical Competence Measure*

Moderator Comparison	<i>z</i>	<i>p</i>	<i>df</i>
Overall			
WJAP & PENS-B	0.63	.529	1,800
WJAP & Panamath	-0.27	.787	1,040
Panamath & PENS-B	-0.79	.427	630
Age Group			
<6 years & 6–9 years			
WJAP	-3.86	<.001	595
Panamath	-0.74	.457	393
Number Type			
Symbolic & Non-Symbolic			
PENS-B	-0.30	.767	222
Panamath	-0.05	.961	406
Presentation Medium			
Computer & Paper			
WJAP	5.47	<.001	515
PENS-B	-0.30	.767	222
Number Range			
0 to 10 & 0 to 20			
Panamath	0.05	.961	406
0 to 20 & 0 to 100			
WJAP	1.25	.212	468
0 to 20 & 0 to 1,000			
WJAP	2.13	.033	537
0 to 100 & 0 to 1,000			
WJAP	0.43	.669	257

Note. WJAP = Woodcock-Johnson III Applied Problems Subtest; PENS-B = Preschool Early Numeracy Screener-Brief; PAE = Percentage Absolute Error.

Presentation Medium

The presentation medium moderation test was found to be statistically significant for the estimation-competence relation for children's performance on the Woodcock-Johnson III Applied Problems subtest ($z[515] = 5.47, p < .001$). However, presentation medium did not significantly moderate the estimation-competence relation for the PENS-B measure. The relation between computer or paper performance on the NLE task and the Woodcock-Johnson III Applied Problems subtest differed in effect size $r = .06$ and $-.40$, respectively. However, the estimation-PENS-B relation was similar across computer and paper presentation mediums, $r(\text{PENS-B}) = -.260$ and $-.222$, respectively.

Number Range

The correlation between performance on the NLE task and the Applied Problems subtest was found to be significantly moderated by the number range that was presented ($z[537] = 2.13, p = .033$). However, all other comparisons were not significant. Children who completed the 0–10 number line and the 0–20 number line demonstrated similar effect sizes for the estimation-competence relation, $r(\text{Panamath}) = -.262$ and $-.267$, respectively. Further, the effect sizes of

the relation between performance on the NLE task and the Woodcock-Johnson III Applied Problems subtest were also similar across 0–20, 0–100, and 0–1,000 number ranges ($r = -.40, -.51, \text{ and } -.55$ respectively).

Extension

To further assess the stability and validity of the NLE task, we also included extension analyses beyond the replication of the study by Schneider and colleagues (2018). Multiple linear regressions were used to examine whether the NLE task was stable over time while controlling for other mathematical measures and whether the NLE task predicted other mathematical measures while controlling for previous time points.

Number Line Estimation Stability

Results from regression analyses assessing the stability of the NLE task are presented in Table 4. Both age and sex were included as covariates. Although the correlation between NLE performance at Time 1 and Time 2 was small to moderate ($r = .24$), when using a regression and controlling for age and sex, children's NLE performance at Time 1 was not a statistically significant predictor of their NLE performance at Time 2 (Model 1 $\beta(SE) = 0.06 (0.04)$, $p = .187$). However, when analyzed as separate studies (Study 1, 6, and 7 in Table 1), results suggested that NLE performance predicted itself in studies 1 ($B = 0.20$ [95% CI: 0.03, 0.38]) and 6 ($B = 0.26$ [95% CI: 0.07, 0.44]), but not 7 ($B = 0.04$ [95% CI: $-0.14, 0.23$]). See Table A1 in the Appendix for more details.

Table 4

Regression Predicting Number Line Estimation Stability

Variable	Model 1			Model 2			Model 3			Model 4		
	<i>B</i> (<i>SE</i>)	β	<i>p</i>	<i>B</i> (<i>SE</i>)	β	<i>p</i>	<i>B</i> (<i>SE</i>)	β	<i>p</i>	<i>B</i> (<i>SE</i>)	β	<i>p</i>
Constant	26.95 (2.29)		< .001	14.58 (2.92)		< .001	30.81 (6.86)		< .001	39.35 (3.82)		< .001
Age	-2.25 (0.32)	-0.37	< .001	2.16 (0.57)	0.30	< .001	-2.56 (1.91)	-0.17	.182	-3.74 (0.69)	-0.38	< .001
Sex	0.99 (0.80)	0.06	.219	2.04 (0.81)	0.14	.013	-0.52 (1.79)	-0.03	.771	0.22 (1.03)	0.01	.835
PAE Time 1	0.06 (0.04)	0.07	.187	-0.03 (0.04)	-0.05	.428	0.05 (0.13)	0.04	.675	0.05 (0.06)	0.05	.374
Mathematical Competence												
WJAP Time 1				-0.76 (0.11)	-0.58	<.001						
PENS-B Time 1							-0.01 (0.20)	-0.01	.965			
Panamath Time 1										-0.07 (0.04)	-0.13	.066
Adjusted R^2	.16			.18			.04			.23		
<i>df</i>	406			290			113			237		

Note. WJAP = Woodcock-Johnson III Applied Problems Subtest; PENS-B = Preschool Early Numeracy Screener-Brief; PAE = Percentage Absolute Error.

When controlling for previous mathematical competence, performance on the NLE task at Time 1 was not a predictor of performance at Time 2 (Model 2 $\beta(SE) = -0.032 (0.040)$, $p = .421$; Model 3 $\beta(SE) = 0.053 (0.125)$, $p = .675$; Model 4 $\beta(SE) = 0.054 (0.060)$, $p = .374$). However, children's performance on the Applied Problems subtest at Time 1 significantly predicted their performance on the NLE task at Time 2, while controlling for prior number line performance (Model 1 $\beta(SE) = -0.691 (0.096)$, $p < .001$). The other mathematical competence measures did not predict performance on the number line task at Time 2.

Number Line Estimation Predictive Validity

Results from the multiple linear regressions are presented in Table 5 by mathematical competence measure. In all regressions, children's age and sex were included as covariates.

Table 5

Regression Predicting Number Line Estimation Predictive Validity

Variable	Model 1			Model 2			Model 3		
	WJAP Time 2			PENS-B Time 2			Panamath Time 2		
	<i>B</i> (<i>SE</i>)	β	<i>p</i>	<i>B</i> (<i>SE</i>)	β	<i>p</i>	<i>B</i> (<i>SE</i>)	β	<i>p</i>
Constant	10.61 (1.17)		< .001	0.07 (3.27)		.982	28.10 (4.78)		< .001
Age	0.08 (0.24)	0.02	.728	2.37 (0.91)	0.24	.010	5.44 (0.85)	0.36	< .001
Sex	-0.69 (0.33)	-0.06	.039	-0.65 (0.85)	-0.06	.448	-2.37 (1.29)	-0.09	.067
PAE Time 1	0.03 (0.02)	0.05	.141	-0.08 (0.06)	-0.09	.206	-0.08 (0.08)	-0.05	.287
Mathematical Competence									
WJAP Time 1	0.73 (0.04)	0.83	< .001						
PENS-B Time 1				0.54 (0.09)	0.52	< .001			
Panamath Time 1							0.37 (0.05)	0.42	< .001
Adjusted <i>R</i> ²		.69			.51			.49	
<i>df</i>		334			113			237	

Note. WJAP = Woodcock-Johnson III Applied Problems Subtest; PENS-B = Preschool Early Numeracy Screener-Brief; PAE = Percentage Absolute Error.

Applied Problems Subtest — The multiple linear regressions suggest the Applied Problems subtest of the Woodcock-Johnson III Tests of Achievement predicted itself when controlling for other variables, $\beta(SE) = 0.727 (0.038)$, $p < .001$. However, the performance on the NLE task did not predict Applied Problems at Time 2 when controlling for Applied Problems Time 1 score, $\beta(SE) = 0.025 (0.017)$, $p = .140$.

Preschool Early Numeracy Screener-Brief — Regression results demonstrated that the PENS-B measure predicted itself when controlling for other variables, $\beta(SE) = 0.537 (0.094)$, $p < .001$. Similar to the Applied Problems subtest, performance on the NLE task was not a statistically significant predictor of PENS-B at Time 2 when controlling for PENS-B Time 1 scores, $\beta(SE) = -0.076 (0.094)$, $p = .206$.

Panamath — Results revealed that Panamath was a statistically significant predictor of itself when controlling for other variables, $\beta(SE) = 0.366 (0.049)$, $p < .001$. Similar to the previous mathematical measures, performance on the NLE task did not predict Panamath at Time 2 when controlling for Panamath Time 1 scores, $\beta(SE) = -0.081 (0.076)$, $p = .287$.

Discussion

The goal of the current study was to conceptually replicate correlational results from a meta-analysis examining the relation between NLE and mathematical competence (Schneider et al., 2018). Results using seven diverse and independent studies demonstrated mixed results. The moderate correlation between the NLE task and mathematical competence replicated. Further, consistent with Schneider et al. (2018), the correlation between the two constructs was moderated by age group, but not in the same direction. However, inconsistent with Schneider et al. (2018), presentation medium and number range also moderated this correlation. Unique to this study was a set of analyses that extended the results of Schneider et al. (2018) by examining the stability and predictive validity of the NLE task while controlling for participant age and sex. Interestingly, these results found that the NLE task was not stable across time, nor was it predictive of later mathematical competence.

Replication

The new findings from the seven independent studies replicated the Schneider et al. (2018) effect sizes and supported the first hypothesis that performance on the NLE task is associated with mathematical competence. In the original study, the overall estimation-competence effect size in the Schneider et al. (2018) meta-analysis was $r = .44$ ($r = -.44$

before recoding). Across the seven studies included in this replication, the strength of the association ranged from $r = -.40$ to $-.35$, and similar to Schneider et al. (2018), the strength of the association remained stable across mathematical competence measures.

Estimation-competence stability across the mathematical competence measures could be due to a few different things. First, the mathematical competence measures are highly correlated. Although, one of the mathematical competence measures was Panamath which is thought to measure a more innate, non-symbolic numerical processing (Halberda et al., 2008) and the Applied Problems and PENS-B measures assess more symbolic, non-innate mathematical skills, they are both mathematical competence measures that are part of the same mathematical construct. Relatedly, the consistency in the NLE-competence relations across mathematical measures could point to the interdependence of these developmental pathways among two foundational processes (non-symbolic and symbolic). Lau and colleagues (2021) found that earlier symbolic number ability was consistently the strongest predictor of approximate number ability. Therefore, although separate constructs, these two processes may work together to better refine individual mathematical skills. Thus, consistency across this relation, regardless of competence measure, may reflect refinement among these processes in the development of early numerical skills.

Results also supported part of our second hypothesis, such that age group moderated the estimation-competence association. The age moderation replicated results from Schneider et al. (2018); however, due to our small sample size of children above 9 years of age, we were unable to assess this third age group that was present in the original meta-analysis. Interestingly, the two age groups we were able to conceptually replicate demonstrated inconsistencies with Schneider et al. (2018), such that children's age group did not moderate the relation in the same way. Our results suggested that the estimation-competence relation was stronger for children in the below 6 years of age group than for children in the 6 to 9 years of age group for both math competence measures available. Schneider et al. (2018), however, found an increasing estimation-competence relation as age group increased. One possible explanation for this could be that children in the below 6 and 6 to 9 age groups did not receive drastically different number ranges in the current replication studies. Instead, these two age groupings received similar number ranges (e.g., 0 to 20) that may be simpler for the 6 to 9-year-olds than those for children below 6. Thus, the number range was better suited for the lower age range and skill level present in the current sample.

Inconsistent with our third hypothesis and findings from Schneider et al. (2018), our results also revealed presentation medium and number range moderated the estimation-competence relation, but only for one mathematical competence measure, the Applied Problems subtest. The estimation-competence relation effect size was closer to what we hypothesized across presentation mediums and the number range for the Panamath measure. This divergent pattern of findings leads us to believe that, perhaps in these samples, the moderating variables were not mutually exclusive and were dependent upon the task variations presented to children. Specifically, the presentation medium and number range samples for the Applied Problems subtest were distinct for both Study 1 and Study 6, such that in Study 6, children were presented with more challenging number line ranges (0–100, 0–1,000) on a computer. In contrast, in Study 1, children were presented with a more straightforward number line range for their age group (0–20) with paper and pencil. Thus, these results may be a function of our samples' restrictions, rather than the moderator itself.

Although complicated, these replication results highlight the necessity for future work to focus on the importance of variation in the NLE task. In their meta-analysis, Schneider et al. (2018) hypothesize that NLE tasks that require fraction estimation strategies would allow for more fine-grained assessments of mathematical knowledge because fraction estimation strategies tend to be more complex than whole number estimation strategies (Rinne, Ye, & Jordan, 2017; Schneider & Siegler, 2010). Their results supported this hypothesis, as the estimation-competence correlation was higher for fractions than it was for whole numbers (Schneider et al., 2018). Our replication results, however, demonstrate a different effect regarding the range of whole numbers presented to children during the NLE task. At some age, the number range 0 to 20 becomes too easy for children, demonstrating a decreasing estimation-competence relation. Therefore, the developmental progression of NLE measurement to match certain age ranges remains unclear. As the number line task exists, it is very difficult to disentangle the effects of age, number range, and other factors from the larger "numerical magnitude" skill that this assessment should measure. It could be that the way in which the NLE task is scored (PAE) is not the most accurate way of measuring numerical magnitude across all developmental age groups as this approach does not consider different strategy use, knowledge of numbers, and so on (Xu, Burr, Douglas,

Susperreguy, & LeFevre, 2021). Taken together, the inconsistency between our replication and the original meta-analysis suggest it is imperative that our field works to have the correct assessment (i.e., number range) fit the participant (i.e., age), and further examine the best way in which to score performance on the NLE task for young children. In sum, our replication efforts have emphasized the importance of utilizing a developmentally appropriate measure to capture numerical magnitude skills.

Extension

In the current replication effort, we also extended the Schneider et al. (2018) meta-analysis by conducting exploratory analyses to assess the stability and predictive validity of the NLE task. Schneider et al. (2018) included results on the predictive validity of the NLE task for later competence and found that it was a moderate predictor (see Table 1 in Schneider et al.). Across all measures of mathematical competence in this study, children's performance on the NLE task at an earlier time point did not predict their performance on the same task at a later time point when controlling for age and sex, or mathematical competence. However, when analyzed as separate studies this finding was inconsistent such that the NLE task predicted itself in two of three individual studies. Interestingly, only one mathematical competence measure predicted children's estimation performance, the Applied Problems subtest. Therefore, there was inconsistency not only in the stability of the NLE task, but also across the mathematical competence measures that predicted the task. These findings speak to the inconsistency in the field of results with the NLE task, and further support our earlier arguments that the numerical magnitude skill cannot be disentangled from other potential moderating variables when using this task.

Consistently, however, our results suggested that the NLE task did not predict any of our three mathematical competence measures while controlling for prior achievement. However, in each case, mathematical competence demonstrated strong stability among all competence measures, even while controlling for children's age at testing, sex, and performance on the number line task. These findings add to our previous results to suggest that the NLE task did not demonstrate strong predictive validity for other mathematical competence measures. Taken together, the inconsistency and lack of evidence in stability, the lack of evidence in predictive validity, and the changing relation with mathematical competence across ages raises into question what, specifically, the NLE task measures.

Although both the stability and predictive validity of the NLE task were exploratory, our results are consistent with other recent studies that have assessed the reliability of the task. For example, multiple studies have shown low internal reliability for various versions of the NLE task (Hawes, Nosworthy, Archibald, & Ansari, 2019; Inglis & Gilmore, 2014; Kolkman, Kroesbergen, & Leseman, 2013). One study examined the stability of the NLE task in a sample with similar age groups and number line ranges as the current replication study (O'Connor, Morsanyi, & McCormack, 2019). Analyses revealed that children's performance on the NLE task was not correlated with later performance, demonstrating that the skills measured by this task may be unstable. Further results suggested the way in which children solve questions on the NLE task may qualitatively change over time (O'Connor et al., 2019). Thus, children's performance on the NLE task may reflect various other skills that develop over time, such as familiarity with numbers (Xu et al., 2021), or strategy use (Xu & LeFevre, 2016), not a single underlying numerical magnitude ability.

Conclusion

In sum, the current study successfully replicated the overall findings from the Schneider et al. (2018) meta-analysis. The strength of the estimation-competence association replicated, as did the finding that the age group of the child was important for the strength of this relation, though not increasingly. Furthermore, consistent with Schneider et al. (2018), the association remained stable across mathematical competence measure and number type. Inconsistent with Schneider et al. (2018), the correlation did not remain stable across presentation medium and number range. Exploratory analyses revealed that the NLE task did not demonstrate strong stability or predictive validity. Thus, our results generally replicated the correlational nature of the NLE task and mathematical competence found in Schneider et al. (2018). However, the current study also highlighted the instability and lack of unique variance provided by the NLE task across seven independent studies. Future research should first focus on understanding the connection between

children's developmental progression and NLE measurement before further investigating the predictive and diagnostic importance of the task for broader mathematical competence.

Funding: Funding for this work came from multiple different grants. Funding for Study 1 comes from Grant 1356118 from the National Science Foundation. Funding for Study 2 and 4 comes from Grant 2016225239 from the Graduate Research Fellowship Program at the National Science Foundation. Funding for Study 5 comes from Grant 1749294 from the CAREER grant at the National Science Foundation. Funding from Study 6 comes from Grant FONDECYT Regular 1180675 from the Chilean National Fund of Scientific and Technological Development (ANID/CONICYT FONDECYT).

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

Competing Interests: The authors have declared that no competing interests exist.

Ethics Approval: All seven datasets had IRB approval and APA ethical standards were followed in completing this research.

Data Availability: For this article, a dataset is freely available (Ellis, Susperreguy, Purpura, & Davis-Kean, 2021).

Supplementary Materials

The Supplementary Materials contain the syntax and a simulated dataset for this study (for access see [Index of Supplementary Materials](#) below).

Index of Supplementary Materials

Ellis, A., Susperreguy, M. I., Purpura, D. J., & Davis-Kean, P. E. (2021). *Supplementary materials to "Conceptual replication and extension of the relation between the number line estimation task and mathematical competence across seven studies"* [Research data and code]. OSF. <https://osf.io/qswav/>

Journal of Numerical Cognition. (Ed.). (2021). *Supplementary materials to "Conceptual replication and extension of the relation between the number line estimation task and mathematical competence across seven studies"* [Open peer-review]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.5228>

References

- Ahmed, S. F., Grammer, J., & Morrison, F. (2021). Cognition in context: Validating group-based executive function assessments in young children. *Journal of Experimental Child Psychology*, 208, Article 105131. <https://doi.org/10.1016/j.jecp.2021.105131>
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, 46(2), 545-551. <https://doi.org/10.1037/a0017887>
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42(1), 189-201. <https://doi.org/10.1037/0012-1649.41.6.189>
- Cipora, K., Patro, K., & Nuerk, H. C. (2015). Are spatial-numerical associations a cornerstone for arithmetic learning? The lack of genuine correlations suggests no. *Mind, Brain, and Education*, 9(4), 190-206. <https://doi.org/10.1111/mbe.12093>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cribbie, R., Beribisky, N., & Alter, U. (2019). *A multi-faceted mess: A review of statistical power analysis in psychology journal articles*. PsyArXiv. <https://doi.org/10.31234/osf.io/3bdfu>
- 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(5880), 1217-1220. <https://doi.org/10.1126/science.1156540>
- del Río, M. F., Susperreguy, M. I., Strasser, K., Cvencek, D., Iturra, C., Gallardo, I., & Meltzoff, A. N. (2020). Early sources of children's math achievement in Chile: The role of parental beliefs and feelings about math. *Early Education and Development*, 32(5), 637-652. <https://doi.org/10.1080/10409289.2020.1799617>

- Ellis, A., Ahmed, S. F., Zeytinoglu, S., Isbell, E., Calkins, S. D., Leerkes, E. M., . . . Davis-Kean, P. E. (2021). Reciprocal associations between executive function and academic achievement: A conceptual replication of Schmitt et al. (2017). *Journal of Numerical Cognition*, 7(3), 453-472. <https://doi.org/10.5964/jnc.7047>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. <https://doi.org/10.3758/BF03193146>
- 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>
- Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L., & Hamlett, C. L. (2014). Sources of individual differences in emerging competence with numeration understanding versus multidigit calculation skill. *Journal of Educational Psychology*, 106(2), 482-498. <https://doi.org/10.1037/a0034444>
- Giner-Sorolla, R., Aberson, C. L., Bostyn, D. H., Carpenter, T., Conrique, B. G., Lewis, N. A., & Soderberg, C. (2019). *Power to detect what? Considerations for planning and evaluating sample size* [Preprint]. Retrieved from <https://osf.io/jnmya/>
- Halberda, J., Mazocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455(7213), 665-668. <https://doi.org/10.1038/nature07246>
- Hawes, Z., Nosworthy, N., Archibald, L., & Ansari, D. (2019). Kindergarten children's symbolic number comparison skills relates to 1st grade mathematics achievement: Evidence from a two-minute paper-and-pencil test. *Learning and Instruction*, 59, 21-33. <https://doi.org/10.1016/j.learninstruc.2018.09.004>
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, 145, 147-155. <https://doi.org/10.1016/j.actpsy.2013.11.009>
- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction*, 25, 95-103. <https://doi.org/10.1016/j.learninstruc.2012.12.001>
- Lau, N. T., Merkley, R., Tremblay, P., Zhang, S., De Jesus, S., & Ansari, D. (2021). Kindergarteners' symbolic number abilities predict nonsymbolic number abilities and math achievement in grade 1. *Developmental Psychology*, 57(4), 471-488. <https://doi.org/10.1037/dev0001158>
- 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, Article 641. <https://doi.org/10.3389/fpsyg.2013.00641>
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 17(5), 714-726. <https://doi.org/10.1111/desc.12152>
- Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). *Bateria III Pruebas de aprovechamiento*. Itasca, IL, USA: Riverside Publishing.
- O'Connor, P. A., Morsanyi, K., & McCormack, T. (2019). The stability of individual differences in basic mathematics-related skills in young children at the start of formal education. *Mind, Brain, and Education*, 13(3), 234-244. <https://doi.org/10.1111/mbe.12190>
- Purpura, D. J., & Logan, J. A. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, 51(12), 1717-1724. <https://doi.org/10.1037/dev0000055>
- Purpura, D. J., Reid, E. E., Eiland, M. D., & Baroody, A. J. (2015). Using a brief preschool early numeracy skills screener to identify young children with mathematics difficulties. *School Psychology Review*, 44(1), 41-59. <https://doi.org/10.17105/SPR44-1.41-59>
- Purpura, D. J., & Simms, V. (2018). Approximate number system development in preschool: What factors predict change? *Cognitive Development*, 45, 31-39. <https://doi.org/10.1016/j.cogdev.2017.11.001>
- Reid, E. E., Baroody, A. J., & Purpura, D. J. (2015). Assessing young children's number magnitude representation: A comparison between novel and conventional tasks. *Journal of Cognition and Development*, 16(5), 759-779. <https://doi.org/10.1080/15248372.2014.920844>
- Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology*, 53(4), 713-730. <https://doi.org/10.1037/dev0000275>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, 89(5), 1467-1484. <https://doi.org/10.1111/cdev.130>
- Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1227-1238. <https://doi.org/10.1037/a0018170>

- Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, 19(3), 341-361. <https://doi.org/10.1111/desc.12395>
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75(2), 428-444. <https://doi.org/10.1111/j.1467-8624.2004.00684.x>
- Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, 62(4), 273-296. <https://doi.org/10.1016/j.cogpsych.2011.03.001>
- Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, 142(1), 193-208. <https://doi.org/10.1037/a0028560>
- Soper, D. S. (2020). *Significance of the difference between two correlations calculator* [Computer software]. Retrieved from <http://www.danielsoper.com/statcalc>
- Susperreguy, M. I., & Davis-Kean, P. E. (2016). Maternal math talk in the home and math skills in preschool children. *Early Education and Development*, 27(6), 841-857. <https://doi.org/10.1111/cdev.12924>
- Thompson, C. A., & Siegler, R. S. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychological Science*, 21(9), 1274-1281. <https://doi.org/10.1177/0956797610378309>
- 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(5), 979-989. <https://doi.org/10.1016/j.neuropsychologia.2013.02.001>
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III NU Complete*. Rolling Meadows, IL, USA: Riverside Publishing.
- Xu, C., Burr, S. D. L., Douglas, H., Susperreguy, M. I., & LeFevre, J. A. (2021). Number line development of Chilean children from preschool to the end of kindergarten. *Journal of Experimental Child Psychology*, 208, Article 105144. <https://doi.org/10.1016/j.jecp.2021.105144>
- Xu, C., & LeFevre, J. A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology*, 52(6), 854-866. <https://doi.org/10.1037/dev0000124>

Appendix

Table A1

Study Breakdown of Regression Predicting Number Line Estimation Stability in Model 1 in Table 3

Variable	Study 1			Study 6			Study 7		
	B (SE)	β	p	B (SE)	β	p	B (SE)	β	p
Constant	3.89 (9.88)		.694	49.42 (7.45)		< .001	30.90 (6.54)		< .001
Age	1.06 (1.65)	0.06	.522	-5.34 (1.04)	-0.48	< .001	-2.61 (1.43)	-0.17	.072
Sex	0.16 (1.24)	0.01	.896	2.68 (1.16)	0.17	.022	-0.50 (1.73)	-0.03	.772
PAE Time 1	0.15 (0.06)	0.20	.022	0.22 (0.08)	0.26	.007	0.05 (0.12)	0.04	.643
Adjusted R ²		.02			.18			.01	
df		139			146			110	

Note. PAE = Percentage Absolute Error.



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



leibniz-psychology.org

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