Journal of Experimental Child Psychology 225 (2023) 105521

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The number line estimation task is a valid tool for assessing mathematical achievement: A population-level study with 6484 Luxembourgish ninth-graders

Sevim Nuraydin [a,*], Johannes Stricker [b], Sonja Ugen [c], Romain Martin [d], Michael Schneider [a]

[a] *University of Trier, D-54286 Trier, Germany*
[b] *Heinrich Heine University Düsseldorf, D-40204 Düsseldorf, Germany*
[c] *University of Luxembourg, L-4365 Esch-sur-Alzette, Luxembourg*
[d] *Ministry of Higher Education and Research, L-2327 Luxembourg, Luxembourg*

ARTICLE INFO

Article history:
Received 22 February 2021
Revised 15 July 2022

Keywords:
Number line
Validity
Fraction
Mathematical achievement
Numerical cognition
Diagnostic

ABSTRACT

The number line estimation task is an often-used measure of numerical magnitude understanding. The task also correlates substantially with broader measures of mathematical achievement. This raises the question of whether the task would be a useful component of mathematical achievement tests and instruments to diagnose dyscalculia or mathematical giftedness and whether a stand-alone version of the task can serve as a short screener for mathematical achievement. Previous studies on the relation between number line estimation accuracy and broader mathematical achievement were limited in that they used relatively small nonrepresentative samples and usually did not account for potentially confounding variables. To close this research gap, we report findings from a population-level study with nearly all Luxembourgish ninth-graders (N = 6484). We used multilevel regressions to test how a standardized mathematical achievement test relates to the accuracy in number line estimation on bounded number lines with whole numbers and fractions. We also investigated how these relations were moderated by classroom characteristics, person characteristics, and trial characteristics. Mathematical achievement and number line estimation accuracy were associated even after controlling for potentially confounding

* Corresponding author.
E-mail address: nuraydin@uni-trier.de (S. Nuraydin).

https://doi.org/10.1016/j.jecp.2022.105521

variables. Subpopulations of students showed meaningful differences in estimation accuracy, which can serve as benchmarks in future studies. Compared with the number line estimation task with whole numbers, the number line estimation task with fractions was more strongly related to mathematical achievement in students across the entire mathematical achievement spectrum. These results show that the number line estimation task is a valid and useful tool for diagnosing and monitoring mathematical achievement.

Introduction

The number line estimation task has gained much attention as a tool for assessing, monitoring, and predicting mathematical achievement (e.g., Booth & Siegler, 2006; Siegler & Booth, 2004). In this task, participants indicate the position of numbers (e.g., whole numbers, fractions) on an empty number line. Only the start point and end point of the line are usually marked and labeled with the respective numbers. The accuracy of the answers is often computed as percent absolute error (PAE) with *PAE* = 100% * (|*correct position* – *estimated position*| / *numerical range of the line*) (Siegler & Booth, 2004). Number line estimation accuracy correlates substantially with measures of broader mathematical achievement for different age groups, task versions, and achievement measures (Schneider et al., 2018).

The number line estimation task has many advantages over other tasks used to assess mathematical achievement. It is easy to administer, needs only brief instructions, can be easily varied in its degree of difficulty, and is understood even by participants with low language competence. The task usually requires less than 5 min because each trial takes only a few seconds and only a few trials suffice to reach substantial correlations with mathematical achievement (e.g., 20 trials; Schneider et al., 2018). The task difficulty can be adapted to different age groups without altering the task itself simply by changing the type of number to be estimated (e.g., whole numbers for younger children, fractions for older children) or the numerical range of the line (e.g., 0–10 for preschool students, 0–1000 for fourth-graders). The task provides unbiased measures of mathematical achievement in that it requires only mathematical knowledge and no knowledge about the world or measurement units (Booth & Siegler, 2006). The task is useful for research because it can be used in multiple settings: the laboratory, the field (e.g., in schools), and even brain imaging settings, for example, functional magnetic resonance imaging (fMRI) scanners (Vogel et al., 2013). The task can be easily presented in a computer-assisted testing environment, and it is easy to record responses in this task electronically.

These advantages suggest that the number line estimation task should be included, among other tasks, in mathematical achievement tests. They also suggest that a stand-alone version of the task could be useful as a brief assessment of mathematical achievement in situations where the time does not permit complete standardized mathematical achievement tests. Complete achievement tests usually comprise many tasks and often take 30–90 min. By contrast, a brief assessment of mathematical competence with the number line task could be useful in large panel studies where many constructs are measured and little time is available per measure. A brief measure could also be useful for keeping the logistical effort manageable in screenings of large groups of children in practical contexts (e.g., schools) to identify students with mathematical giftedness or mathematical learning disabilities. Finally, brief assessments of mathematical achievement could be useful in field studies in schools when practical constraints do not allow for a longer test time. Obviously, a short assessment like the number line estimation task could never fully replace a standardized mathematical achievement test. A short assessment will generally have lower reliability and validity than a more comprehensive test. However, there are many situations in which a full assessment is not possible, and in these situations using a short quantitative assessment is better than using no quantitative assessment at all.

The number line estimation task might also be useful for diagnosing mathematical giftedness or dyscalculia because measuring mathematical achievement is a central component of the diagnostic process for mathematical giftedness or dyscalculia. Lüftenegger et al. (2015) wrote, "Mathematically gifted students are often identified with standardized achievement tests (Sheffield, 1994; Ziegler & Stoeger, 2010). To categorize people as gifted, different cutoff points ranging from the 80th to the 99th percentile have been established" (p. 228). Likewise, to diagnose dyscalculia, researchers usually show that a student has mathematical achievement far below the average that cannot be explained by domain-general cognitive impairments (e.g., low intelligence). Previous studies have used cutoff values ranging from very strict (e.g., below the first percentile; Landerl et al., 2004) to very broad (e.g., below the 35th percentile; Jordan et al., 2003). When screening populations for giftedness or dyscalculia, short screening instruments need to be used due to the many persons to be tested (Gersten et al., 2011). The number line estimation task might be useful for this.

Relations between number line estimation and mathematical achievement

Previous studies have concluded that the number line estimation task might be useful for measuring aspects of mathematical achievement. A meta-analysis of 263 effect sizes from studies with 10,576 children found an average correlation of $r^+ = .443$ between number line estimation accuracy and measures of mathematical achievement (Schneider et al., 2018). The correlation was moderate ($.351 \leq r^+ \leq .536$) and statistically significant for paper-and-pencil versions and computerized versions of the number line estimation task and for different mathematical competence measures, including assessments of counting, mental arithmetic, written arithmetic, school grades for mathematics, and standardized mathematical achievement tests. Moderator analyses showed that the average correlation between number line estimation accuracy and mathematical competence was highest in students older than 9 years ($r^+ = .491$) and lowest in students younger than 6 years ($r^+ = .296$). In addition, the correlation was higher for number line estimation with fractions ($r^+ = .523$) than for number line estimation with whole numbers ($r^+ = .409$).

Several studies have investigated longitudinal relations between number line estimation accuracy and mathematical competence (e.g., Bailey et al., 2014; Jordan et al., 2013). Meta-analytic aggregation of these findings showed that the correlation is high when the two constructs were measured simultaneously (191 effect sizes, $r^+ = .427$), when the number line estimation task was used to predict mathematical achievement over time (33 effect sizes, $r^+ = .496$), or when mathematical achievement was used to predict number line estimation accuracy over time (39 effect sizes, $r^+ = .538$) (Schneider et al., 2018).

Broad agreement exists that the number line estimation task is closely related to mathematical achievement because it assesses a central component of mathematical competence. However, there is an ongoing debate about what this component is. According to one view (Siegler & Opfer, 2003), number line estimation accuracy reflects how well learners mentally represent numerical magnitudes in the approximate number system (i.e., the so-called mental number line), which has been shown to support the acquisition of mathematical achievement (Dehaene et al., 2008; Yuan et al., 2020). An alternative view assumes that number line estimation mainly requires proportional reasoning to relate the start point of the line, the end point of the line, and the position of the number to be estimated (Barth & Paladino, 2011; Zax et al., 2019). Proportional reasoning is a form of relational reasoning that is relevant in a wide range of mathematical tasks and might also facilitate the further acquisition of mathematical achievement (Gouet et al., 2020). The number line estimation task might also assess visuospatial skills (Gunderson et al., 2012), counting competence (Petitto, 1990), intelligence (Schneider et al., 2009), and socioeconomic status (SES) (Ramani & Siegler, 2008) to some extent. None of these views excludes the others. Proportional reasoning operates on numerical magnitude representations and requires reasoning (i.e., fluid intelligence). Visuospatial skills and counting of segments on the line might help in mapping proportion estimates onto the number lines. In our view, it is likely that the number line estimation task assesses a mix of interrelated mathematical and nonmathematical competencies, all of which contribute to mathematical thinking and problem solving. Therefore, in the current study, we analyzed how number line estimation accuracy relates to a broad range of trial, person, and classroom characteristics.

The need for large-scale studies controlling for potentially confounding variables

Findings on the relation of number line estimation accuracy with mathematical achievement are limited in at least three ways. One limitation is that most previous studies have used relatively small convenience samples, so the generalizability of the findings to various subpopulations is unclear. For example, it is unclear whether the relation between number line estimation accuracy and mathematical achievement differs between genders, between immigrant and nonimmigrant students, or between students on different educational tracks. In addition, thus far, a representative large-scale sample is missing that allows deriving benchmarks for the average number line estimation accuracy in different subpopulations. These benchmarks could subsequently help to interpret study results and to identify students with special education needs (cf. Geary et al., 2008). For example, when a ninth-grade student has a PAE of 6% when estimating a whole number on a number line from 0 – 1000, is this performance low or high compared with the population? Is this person in the highest or lowest quartile of the accuracy distribution? Is the error score of 6% so low that it indicates near-perfect estimation performance? If this person is a boy with a migration background who was not tested in his native language, how did he perform compared with other boys with a migration background? Benchmarking data could help to answer such questions.

A second limitation of previous work using the number line estimation task is that different task versions might be differentially suited for assessing mathematical achievement across the mathematical achievement spectrum. For example, number line estimation with whole numbers might be well-suited to identify dyscalculic students, but this task version might produce ceiling effects in high-ability students. Conversely, number line estimation with fractions might have sizable discriminative power in the upper mathematical achievement spectrum but might be unsuitable for identifying dyscalculic students due to floor effects in below-average mathematics students.

Third, thus far, most studies and the meta-analysis (Schneider et al., 2018) have focused on bivariate relations. However, number line estimation accuracy and mathematical achievement may be influenced by confounding variables such as test anxiety and motivation that might cause a partly spurious correlation between the two variables. Three studies investigating unique relations of number line estimation accuracy and mathematical achievement found robust relations after controlling for cognitive and demographic variables (e.g., Bailey et al., 2014; Geary, 2011; Jordan et al., 2013). For example, accuracy in the whole-numbers number line estimation task predicted later fraction magnitude understanding after controlling for whole number arithmetic proficiency, domain-general cognitive abilities, parental income, education, race, and gender (Bailey et al., 2014). However, thus far, no study has investigated the relation of mathematical achievement after simultaneously controlling for classroom-level characteristics, person-level characteristics, and trial-level characteristics. It is currently unclear to what degree the correlation between number line estimation accuracy and mathematical achievement decreases when controlled for a wide range of potentially confounding variables in a nationally representative sample.

Source of individual differences in number line estimation accuracy

In our study, we also investigated predictors of number line estimation performance. These predictors indicate possible sources of individual differences in number line estimation performance and are relevant when the task is used for diagnostic purposes. The dataset used in the current study allowed us to investigate predictors on the trial-level characteristics, person-level characteristics, and classroom-level characteristics. We introduce these potential predictors in the following sections.

Classroom characteristics

Educational tracks are used in many countries to group students based on their (actual or perceived) aptitude. The tracks tend to differ in their average achievement. If students on the highest achieving track also receive more or better instruction on solving number line estimation tasks, this would create a spurious correlation between number line estimation and mathematical achievement. We know of no study investigating this question.

Person characteristics

Mathematical anxiety (i.e., feeling of tension, worry, and fear in situations involving math-related activities; Wang et al., 2015) is negatively related to mathematical achievement (for a meta-analysis, see Namkung et al., 2019). Number line estimation tasks constitute testing situations involving numerical values in which mathematical anxiety may impair performance. Thus far, empirical findings on the relation of number line estimation accuracy with mathematical anxiety are mixed. Some findings found a negative relation between mathematics anxiety and number line estimation accuracy (e.g., Hart et al., 2016), whereas other studies found no relation (e.g., Kucian et al., 2018).

Mathematical self-concept (i.e., the perception of one's capacity and performance level in mathematics) is related to mathematical achievement (Parker et al., 2014). Thus far, it is unclear whether the mathematical self-concept is also related to the number line estimation accuracy as an indicator of basic numerical competencies.

Test motivation (i.e., the willingness to invest effort and persistence into a test) is related to performance in various tests of mathematical competence and cognitive ability (e.g., Penk & Richter, 2017). Whereas robust evidence shows that motivational factors are related to academic achievement (Kriegbaum et al., 2018), to date the relation between test motivation and number line estimation accuracy is unclear.

Conscientiousness (i.e., the tendency to be reliable, organized, and self-disciplined; Costa & McCrae, 1992) is related to mathematical achievement (Alcock et al., 2014). The relation of conscientiousness with number line estimation accuracy thus far has not been investigated. On the one hand, it seems plausible that conscientiousness is positively related to number line estimation accuracy because number line estimation accuracy substantially overlaps with mathematical achievement. On the other hand, the number line estimation task assesses relatively basal mathematical competencies, which might be less strongly influenced by personality than more complex competencies.

Higher SES is associated with higher mathematical achievement (Agirdag et al., 2012). Students from low-income households often start school with lower mathematical competence levels and show slower growth in mathematical competencies than students from high-income households (Jordan & Levine, 2009). There is some evidence that students with low socioeconomic backgrounds also have larger difficulties with numerical magnitude understanding (Ramani & Siegler, 2008). However, to date, an empirical test of the SES–number line estimation accuracy relation in a large and representative sample is missing.

Some empirical studies found no or negligible gender differences in number line estimation accuracy (e.g., Reinert et al., 2017), leading to the conclusion that "sex differences in mathematical ability are unlikely to be related to number line estimation" (Tosto et al., 2018, p. 797). Other studies found evidence for gender differences in number line estimation (e.g., Gunderson et al., 2012). To date, it is unclear to what extent differences in motivational variables (e.g., mathematical self-concept) account for these gender differences.

Higher proportions of migrant students in a classroom are associated with lower levels of mathematical achievement, although control variables mitigate this relation (Agirdag et al., 2012). It is unclear whether migration status is also related to number line estimation accuracy, for example, due to language problems during classroom instruction or language effects on number magnitude representations (e.g., Helmreich et al., 2011).

In previous studies, participants completed the number line estimation task on screens with varying resolutions, for example, ranging from 1024×768 pixels (Torbeyns et al., 2015) to 1920×1200 pixels (Reinert et al., 2017). On screens with a higher resolution, it might be easier to indicate the correct position of a number (Cohen & Ray, 2020). Thus, the screen resolution might partly account for individual differences in number line estimation accuracy when using computer-based presentations.

Trial characteristics

A better understanding of how trial characteristics influence number line estimation accuracy can inform the design of standardized assessments using this task. Previous studies found relatively large differences in number line estimation accuracy between trials and identified several characteristics of the presented numbers that explain part of this variance (Kodosh et al., 2008; Schneider et al., 2008; Siegler et al., 2011). We included these number characteristics as predictors in our model to compare how much variance of number line estimation accuracy is explained by classroom-level characteristics, person-level characteristics, and trial-level characteristics, respectively. In previous studies, number line estimation accuracy tended to be lower for fractions than for whole numbers due to the greater complexity of fractions (Siegler et al., 2011). The accuracy was higher around orientation points (i.e., the start point, the midpoint, and the end point of the line; Schneider et al., 2008; Sullivan et al., 2011). It decreased with increasing number size, which could indicate that larger numbers are more difficult to process than smaller numbers (size effect; Kodosh et al., 2008).

The current study

We aimed to contribute to the validation of the number line estimation task as a tool for diagnosing and monitoring mathematical achievement. To this end, we examined the correlation between number line estimation accuracy before and after controlling for potential confounds as well as sources of individual differences in number line estimation accuracy. We included the number line estimation task in the Luxembourgish Épreuves Standardisées (ÉpStan) large-scale study for monitoring students' school achievement in the 2012–2013 school year (Martin et al., 2015). ÉpStan aims at evaluating the Luxembourgish national educational system by testing to what extent primary and secondary school students reached educational standards defined by the Luxembourg Ministry of Education and Vocational Training. The study included standardized achievement tests for the school subjects of mathematics, French, and German as well as questionnaires about further person characteristics. We were given the opportunity to also present about 19 trials of the number line estimation task to the ninth-graders participating in this study.

This design allowed us to investigate three research questions. First, how high is the PAE for whole-number estimation and fraction estimation in the overall population of Luxemburgish ninth-graders and in relevant subpopulations (i.e., the three education tracks), students in different quartiles of achievement, girls and boys, and students with and without migration status (Research Question 1)? These results can serve as benchmarks to evaluate a sample's or an individual's number line estimation performance compared with population data in future studies.

Second, how strongly does number line estimation predict mathematical achievement before and after controlling for person characteristics and classroom characteristics (Research Question 2)? The answer to this question indicates to what extent we can replicate the bivariate correlations between number line estimation and mathematical achievement found in a previous meta-analysis (Schneider et al., 2018). Going beyond the meta-analysis, here we examined to what extent the relation weakens when it is controlled for person characteristics and classroom characteristics that might have influenced estimation accuracy as well as achievement.

Finally, what variance components of number line estimation PAE are explained by classroom-level characteristics (educational track), person-level characteristics (e.g., math anxiety), and trial characteristics (e.g., the distance of the correct position from the nearest orientation point on the line), respectively (Research Question 3)? These findings can broaden our understanding of what the number line estimation task actually measures. When number line estimation strongly covaries with mathematical competence, this would support the convergent validity of the number line task as a measure of mathematical achievement. When number line estimation weakly covaries with other variables (e.g., test motivation, conscientiousness, mathematical anxiety), this would support the discriminant validity of the number line task as a measure of mathematical achievement.

Method

Participants

A total of 6484 ninth-graders from 348 classrooms in 33 schools in Luxembourg participated in the ÉpStan study (49.0% girls, 50.4% boys, and 0.6% no gender information). The sample was nationally representative in that it comprised nearly all ninth-graders in Luxembourg. The sample mean age was 14 years (*SD* = 0.91), ranging from 12 to 19 years. The majority of participating students had Luxembourgish citizenship (59.2%), followed by Portuguese (24.7%), French (2.2%), Montenegrin (1.8%), and Serbian (1.6%) citizenship. One half the sample reported a migration background (51.8%) as either first-generation immigrants (22.3%) or second-generation immigrants (29.5%).

Students on all three tracks of the Luxembourgish three-track secondary school system participated in this study. The three tracks are the Enseignement Secondaire (ES; classical secondary education; n = 1837), the Enseignement Secondaire Technique (EST; general secondary education; n = 3848), and the Enseignement Préparatoire (EST-PRE; vocational training; n = 799). In Luxemburg, all students attend primary school together for 6 years, after which they are directed to schools that are on one of the three tracks. Schools on the ES level are for the strongest students and prepare them for studying at universities. Schools on the EST level are for medium-strong students and allow them to subsequently study at universities or start vocational training. Students attend schools on the EST-PRE during vocational training. The names of the tracks have changed since our data collection. We use the original names here.

Procedure

The school monitoring program is anchored in the Luxembourgish national law and has been approved by the national committee for data protection. The participants and their legal guardians were informed before the data collection and had the possibility to opt out. We administered all tests in the computer rooms of the students' schools using LUCET's (Luxembourg Center for Educational Training) web-based Online Assessment System (OASYS). The achievement measures were computer-based and comprised a standardized test of mathematical achievement and two standardized language tests (German and French). For each classroom, the study was conducted in a group setting with 15-min breaks provided between the three tests. For the mathematical achievement test and the number line estimation task, students could choose between French or German as the test language. Students could also change the language during the test. For each of the three achievement tests, students were given 50 min. The test order was randomized between participants. The students completed the number line estimation task immediately before the mathematical achievement test. Finally, the participating students completed a self-report questionnaire covering demographic and motivational characteristics, which took approximately 25 min.

Measures

Number line estimation

The number line estimation task was computerized and comprised 19 trials presented in a fixed order. The relatively small number of trials was due to the typically very limited time for additional measures in large-scale student achievement studies. In the first 10 trials, the students indicated the position of ten whole numbers (450, 946, 731, 500, 129, 301, 677, 48, 865, and 271) on a number line ranging from 0 to 1000. In the next 9 trials, the participating students indicated the position of nine fractions (1/2, 4/9, 2/7, 7/9, 1/9, 1/1, 1/3, 3/5, 4/5) on a number line ranging from 0 to 1. In 4 trials the fraction magnitude could be written as a decimal with a finite number of digits (1/2, 3/5, 4/5, and 1/1), and in 5 trials the fraction magnitude could be written as a decimal with an infinite number of digits (1/9, 2/7, 1/3, 4/9, and 7/9). Only the end points of the line were marked and labeled (with 0 and 1000 for whole numbers and with 0 and 1 for fractions). There were no additional marks or labels on the number line. The screen resolution ranged from 853 $\times$ 683 pixels to 1787 $\times$ 1005 pixels. On all screens, the number line was 800 pixels long and had a line width of 8 pixels. On the 0–1000 number lines, each numerical unit (e.g., the distance between 0 and 1) was 800/1000 = 0.8 pixels wide. This explains why the PAE is rarely exactly 0%, not even for students with perfect number knowledge (cf. Cohen & Ray, 2020).

The participants entered their answer by moving a slider to a position on the line and confirming their answer by pressing a button. The slider was located completely at the left on the number line at the start of each trial. The participants could go back to previous trials and change their answers when they wanted. We computed estimation accuracy as *PAE* = 100% * (|*correct position* − *estimated position* | / *numerical range*). Thus, a higher PAE indicates lower estimation accuracy. The average total duration for the 19 trials was 4.03 min (*SD* = 2.05), and the mean solution time per trial was 12.88 s (*SD* = 6.62). The internal consistencies of both task versions were good (Cronbach's α = .84 for whole numbers and α = .79 for fractions).

Mathematical achievement

We assessed mathematical achievement with a standardized test covering mathematical competencies in (a) numbers and operations, (b) space and form, and (c) functions and change (MENFP [Ministère de l'Education nationale et de la Formation professionnelle], 2008). The test was constructed by a team of researchers, active teachers, and curriculum experts. The newly developed items were pretested (for Grade 9 in a sample of 263 students) and evaluated based on psychometric criteria. The mathematical problems were similar to the ones used in other large-scale student achievement studies, for example, PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study). Example items have been published online (LUCET, n.d.). The test included problems with a wide range of difficulties. Most problem statements described typical everyday life situations (e.g., selling candy, water dripping from a crane) and included figures, tables, or similar materials. One half the items had an open item format (e.g., correct number to be typed in), and the other half had a closed item format (e.g., multiple choice). The test had an average duration of 41.18 min (*SD* = 10.33). The test score was standardized to a mean of 500 (*SD* = 100). A detailed description of how the tests were constructed, scored, scaled, and validated can be found in the EpStan technical report (Fischbach et al., 2014).

Three different achievement test versions were used to account for the assumed competency level of each of the three Luxembourgish secondary educational tracks (ES, EST, and EST-PREP). The number of tasks differed for the three test versions (ES and EST: 28 tasks; EST-PRE: 24 tasks). All three test versions contained different proportions of easy, medium, and difficult items, but at least one third of the items were presented for all school forms (Fischbach et al., 2014). These items served as anchor items. The items were scaled with a unidimensional Rasch model. Thus, the three test versions were brought to the same scale and could be analyzed together (Fischbach et al., 2014).

Mathematical anxiety

Mathematical anxiety was assessed with 3 items (e.g., "I am very nervous before mathematics exams"; Fischbach et al., 2014). The students rated each item on a 4-point Likert scale ranging from 1 (*not applicable*) to 4 (*applicable*). The scale had been empirically examined in a previous study (Gogol et al., 2014). This study found an acceptable internal consistency (α = .84) as well as plausible interrelations with other measures of anxiety, self-concept, achievement, and other student characteristics.

Mathematical self-concept

Mathematical self-concept was measured by 3 items translated and adapted from the Self-Description Questionnaire (SDQ; e.g., Marsh & O'Neill, 1984), which is widely used today. An example item is, "I am good at mathematics" (Fischbach et al., 2014). The participating students rated each item on a 4-point Likert scale ranging from 1 (*does not apply*) to 4 (*does apply*). The adapted items had an acceptable internal consistency (α = .84) and plausible interrelations with other measures of self-concept, math anxiety, achievement, and other student characteristics (e.g., Brunner et al., 2010; Gogol et al., 2014; Keller et al., 2016).

Test motivation

The participating students rated their test motivation on 1 item with a 4-point Likert scale ranging from 1 (*not motivated at all*) to 4 (*highly motivated*).

Conscientiousness

Conscientiousness was assessed with 4 items (e.g., "I work orderly"; Fischbach et al., 2014). The participating students rated each item on a 4-point Likert scale ranging from 1 (*not applicable*) to 4 (*applicable*). The scale demonstrated desirable characteristics, including an acceptable internal consistency (α = .75), in previous studies (e.g., Keller et al., 2016).

Demographic characteristics

The participating students indicated their age, gender, migration status, and parents' profession. The information on nationality stems from another official database. We assessed students' SES by coding parents' professions according to the International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom et al., 1992). In the statistical analyses, we used the highest SES of a student's parent.

Trial characteristics

For each trial, we included the number value, the denominator value in trials with fractions, the distance of the nearest orientation point, the position in the presentation order, whether the fraction could be written as a decimal with finite or infinite digits of numbers, and the sum of the presented digits in our analysis. We defined the labeled start point, the midpoint, and the labeled end point of the number line as orientation points and operationalized the nearest distance of each stimulus to one of these three orientation points as the trial's distance to an orientation point. For both number types, we calculated the sum of the presented digits to investigate whether numbers with higher digits are generally more difficult to estimate than numbers with smaller digits. For example, the sum of digits for the presented whole number 271 was 10, and the sum of digits for the presented fraction 1/2 was 3.

Statistical analyses

Because we used population-level data, we focused on descriptive results and standardized effect sizes (e.g., standardized regression weights). We report *p* values for significance tests as ancillary information. Significance tests are used to infer population characteristics from sample characteristics. They are only of limited usefulness when analyzing population-level data (Cowger, 1984; Gigerenzer et al., 2004).

We investigated how strongly number line estimation predicts mathematical achievement (Research Question 2) by conducting two-level regression analyses with persons on Level 1 and classrooms on Level 2. Many schools had only one classroom for each grade, so we could not model schools as a third level. Whenever the educational track was used as a predictor, it was dummy-coded with the middle track of the three Luxembourgish educational tracks (EST) as the reference category. We first used bivariate regression models separately for whole-number and fraction estimation to analyze how strongly number line estimation is associated with achievement. We then entered all other person and classroom characteristics into the models and conducted multiple regressions to examine whether number line estimation predicts achievement over and above the other variables. In addition, we examined classification accuracy by analyzing to what extent students' quartile in the number line estimation accuracy distribution was the same as their quartile in the mathematical achievement distribution.

We investigated to what extent variance in number line estimation accuracy can be explained by classroom characteristics, person characteristics, and trial characteristics by using three-level regression models with trials on Level 1, persons on Level 2, classrooms on Level 3, and number line estimation accuracy as the criterion. We investigated each predictor in a bivariate regression and then investigated the combined influence of all predictors in a multiple regression.

The multilevel regressions were conducted with Mplus (Muthén & Muthén, 1998–2017). We accounted for the non-normal distribution of number line estimation PAE by using the robust maximum likelihood estimator (MLR).

Results

Preliminary analyses

Three-level baseline regression models revealed that the largest proportion of variance in number line estimation PAE was on the trial level (67.1% for whole numbers, 70.9% for fractions), followed by the person level (26.2% for whole numbers, 16.9% for fractions) and the classroom level (6.7% for whole numbers, 12.2% for fractions). So, the intraclass correlation coefficients (ICCs) for trials nested within persons were .262 for whole numbers and .169 for fractions. The ICCs for persons nested within classrooms were .067 for whole numbers and .122 for fractions.

Research Question 1: Mean estimation accuracy in the population and subpopulations

In line with previous studies, students estimated whole numbers (mean PAE = 5.25%) more accurately than fractions (mean PAE = 11.74%). The histograms in Figs. 1 and 2 indicate that the PAEs for whole numbers and fractions were right-skewed. Skewness and kurtosis (see Table S1 in online supplementary material) were below the acceptable cutoff (i.e., ±2; Gravetter & Wallnau, 2014) for fractions but not for whole numbers. See Table S1 for further breakdowns of PAE by the educational track, quartile of the achievement distribution, gender, and migration background. As expected, PAE was generally lower for students on higher educational tracks and in higher achievement quartiles. Figs. S1 to S14 in the supplementary material display the histograms for the four mathematical achievement quartiles and the three educational tracks. PAE was lower for boys than for girls and was lower for students without versus with migration background (see Table S1).

Research Question 2: Number line estimation accuracy as predictor of mathematical achievement

We conducted two-level regression analyses with students on Level 1 and classrooms on Level 2 to examine how strongly number line estimation accuracy predicted mathematical achievement (see Table 1). The ICC for students nested within classes was .477. In a bivariate regression, accuracy in number line estimation with whole numbers had a standardized regression weight of $\beta = -.225$ and explained a variance proportion of $R^2 = .051$ of mathematical achievement. The sign is negative because PAE is an error score. We then also included all other person characteristics (mathematical self-concept, math anxiety, test motivation, conscientiousness, gender, SES, migration status, and computer screen diagonal) and classroom characteristics (educational track) as predictors and conducted a multiple regression. This weakened the relation between number line estimation and achievement ($\beta = -.154$). The predictors together explained variance proportions of $R^2 = .193$ on the person level and of $R^2 = .791$ on the school level. The proportion of variance explained on the school level is so high because students in different tracks of the educational system differ strongly in their achievement and we used the educational track as the classroom-level predictor of achievement.

We repeated these analyses for number line estimation with fractions (see Table 1). In the bivariate regression, number line estimation with fractions had a standardized regression coefficient of $\beta = -.384$ and explained a variance proportion of $R^2 = .147$. In the multiple regression, this relation was slightly weaker ($\beta = -.264$). In the multiple regression, all predictors together explained variance proportions of $R^2 = .232$ on the person level and $R^2 = .787$ on the school level.

We also examined the classification accuracy, that is, to what extent students' quartile in the distribution of number line estimation PAE can be used to predict their quartile in the distribution of

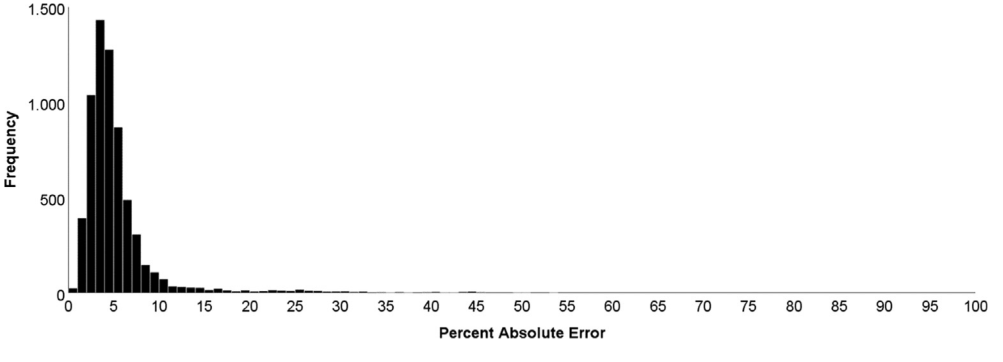

Fig. 1. Distribution of percent absolute error for whole-number estimation.

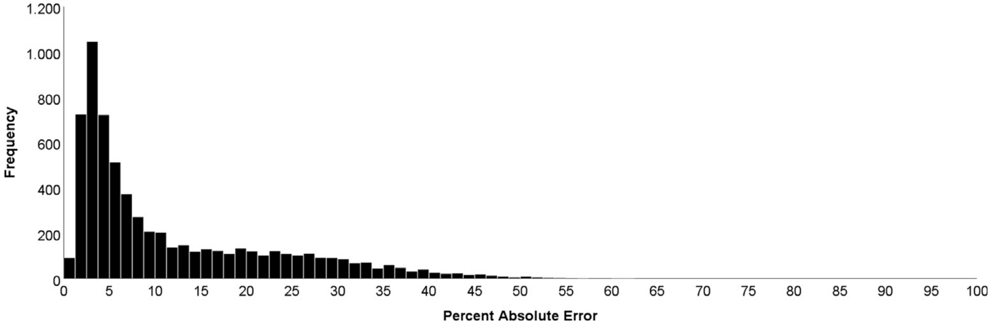

Fig. 2. Distribution of percent absolute error for fraction estimation.

Table 1
Standardized regression weights (β) and proportions of explained variance (R^2) from two-level regressions with classroom- and person-level characteristics as predictors of mathematical achievement.

Predictor	Mathematical achievement					
	Bivariate regression		Multiple regression			
			Number line estimation with whole numbers		Number line estimation with fractions	
	B	R^2	β	R^2	β	R^2
Level 2: Classrooms				.791[b*]		.787[b*]
Educational track (0 = EST, 1 = ES)	.782*	.612[a*]	.669*	.421[c]	.679*	.434[c]
Educational track (0 = EST, 1 = EST-PREP)	−.595*	.355[a*]	−.443*	.176[c]	−.429*	.155[c]
Level 1: Persons				.193[b*]		.232[b*]
Whole-number estimation accuracy	−.225*	.051[a*]	−.154*	.021[c]	–	–
Fraction estimation accuracy	−.384*	.147[a*]	–	–	−.264*	.056[c]
Mathematical self-concept	.241*	.058[a*]	.258*	.038[c]	.223*	.024[c]
Math anxiety	−.208*	.043[a*]	−.106*	.009[c]	−.098*	.008[c]
Test motivation	.130*	.017[a*]	.044*	.006[c]	.043*	.004[c]
Conscientiousness	.083*	.007[a*]	−.042*	.001[c]	−.026	.000[cd]
Gender (1 = female, 2 = male)	.204*	.042[a*]	.098*	.014[c]	.079*	.009[c]
Socioeconomic index	.252*	.063[a*]	.005	.012[c]	.001	.006[c]
Migration status (0 = yes, 1 = no)	.212*	.045[a*]	.070*	.004[c]	.061*	.003[c]
Computer screen diagonal in pixels	.019	.000[a]	.030	.001[c]	.035	.001[c]

Note. ES, Enseignement Secondaire (highest track of the three-track Luxembourgish school system); EST, Enseignement Secondaire Technique (middle track of the three-track Luxembourgish school system); EST-PREP, Enseignement Préparatoire (lowest track of the three-track Luxembourgish school system).
[a] R^2 refers to the unique variance explained by the included predictor in the bivariate regression analyses.
[b] R^2 refers to the total variance explained by the set of predictors on each level.
[c] R^2 for each predictor was calculated as the difference between the total value of R^2 in the multiple regression including all predictors and the total value of R^2 after running the same model without the relevant predictor.
[d] Values less than zero were winsorized to zero.
[*] $p < .05$.

mathematical achievement. The results are shown in the supplementary material (Table S2 for whole-number estimation and Table S3 for fraction estimation). For whole-number estimation, the estimation accuracy quartile and the mathematical achievement quartile were the same for 38.9% of the students. They diverged by only one quartile (e.g., a person is in the third quartile in one distribution and in the fourth quartile in the other) for another 40.3% of the students. For fraction estimation, the estimation accuracy quartile and the mathematical achievement quartile were the same for 43.5% of the students. They diverged by only one quartile for another 41.7% of the students.

Research Question 3: Correlates of number line estimation accuracy on the classroom, person, and trial levels

Table 2 displays the results from a series of three-level bivariate regressions with classroom-level, person-level, and trial-level variables as predictors of the PAE in the number line estimation tasks. In the whole-number estimation task, students on the lowest educational track had a higher PAE (β = .687) and students in the highest educational track had a lower PAE (β = −.539) than students in the middle educational track. Among all variables on the person level, mathematical achievement was most strongly related to number line estimation PAE (β = −.377). Relations with other motivational and demographic variables were considerably weaker (all βs ≤ |.168|). On the trial level, all included characteristics were related to the PAE in the whole-numbers number line estimation task. The strength of these bivariate relations ranged from β = .289 for the distance from the nearest orientation point to β = .076 for the number value. When all predictors were entered as predictors in a multiple regression, mathematical achievement predicted number line estimation PAE over and above the other predictors with β = −.248. In general, the effect sizes found in the bivariate regressions were higher than the ones found in the multiple regressions. This was expected because all the predictors

Table 2

Standardized regression weights (β) and proportions of explained variance (R^2) from three-level regressions with classroom-level, person-level, and trial-level characteristics as predictors of number line estimation percent absolute error.

Predictor	Whole-number estimation accuracy				Fraction estimation accuracy			
	Bivariate regression		Multiple regression		Bivariate regression		Multiple regression	
	B	R^2	β	R^2	β	R^2	β	R^2
Level 3: Classrooms				.465[b*]				.614[b*]
Educational track (0 = EST, 1 = ES)	−.539*	.290[a*]	−.203*	.049[c]	−.617*	.381[a*]	−.312*	.064[c]
Educational track (0 = EST, 1 = EST-PREP)	.687*	.472[a*]	.602*	.358[c]	.681*	.464[a*]	.645*	.396[c]
Level 2: Persons				.098[b*]				.288[b*]
Mathematical achievement	−.377*	.142[a*]	−.248*	.000[c]	−.564*	.318[a*]	−.411*	.037[c]
Mathematical self-concept	−.153*	.023[a*]	−.042	.008[c]	−.284*	.081[a*]	−.143*	.018[c]
Math anxiety	.134*	.018[a*]	.028	.021[c]	.209*	.044[a*]	.021	.002[c]
Test motivation	−.084*	.007[a*]	−.019	.034[c]	−.118*	.014[a*]	−.004	.017[c]
Conscientiousness	.002	.000[a]	.017	.001[c]	.040	.002[a]	.066*	.004[c]
Gender (1 = female, 2 = male)	−.168*	.028[a*]	−.107*	.068[c]	−.263*	.069[a*]	−.145*	.053[c]
Socioeconomic index	−.024	.001[a]	.022	.001[c]	−.056*	.003[a]	−.025	.000[cd]
Migration status (0 = yes, 1 = no)	−.048*	.002[a]	−.005	.000[cd]	−.084*	.007[a*]	−.044*	.006[c]
Computer screen diagonal in pixels	−.049	.002[a]	−.037	.023[c]	−.003	.000[a]	−.021	.000[c]
Level 1: Trials				.109[b*]				.037[b*]
Number value	.076*	.006[a*]	.012	.019[c]	−.007	.000[a]	.101*	.002[c]
Denominator value	–	–	–	–	−.051*	.003[a*]	.429*	.004[c]
Distance from nearest orientation point	.289*	.083[a*]	.279*	.080[c]	.111*	.012[a*]	.029*	.001[c]
Position in the presentation order	.161*	.026[a*]	.041*	.021[c]	.101*	.010[a*]	.133*	.008[c]
Finite number of digits in decimal notation (0 = no, 1 = yes)	–	–	–	–	−.029*	.001[a*]	−.006	.000[c]
Sum of digits	.147*	.022[a*]	.120*	.027[c]	.113*	.013[a*]	−.263*	.002[c]

Note. ES, Enseignement Secondaire (highest track of the three-track Luxembourgish school system); EST, Enseignement Secondaire Technique (middle track of the three-track Luxembourgish school system); EST-PREP, Enseignement Préparatoire (lowest track of the three-track Luxembourgish school system).

[a] R^2 refers to the unique variance explained by the included predictor in the bivariate regression analysis.

[b] R^2 refers to the total variance explained by the set of predictors on each level.

[c] R^2 for each predictor was calculated as the difference between the total value of R^2 in the multiple regression including all predictors and the total value of R^2 after running the same model without the relevant predictor.

[d] Values less than zero were winsorized to zero.

[*] $p < .05$.

were relevant for mathematics learning and hence were intercorrelated. Due to this multicollinearity, the results of the multiple regressions need to be interpreted with caution.

In the fraction estimation task, students in the lowest educational track displayed a higher PAE (β = .681) and students in the highest educational track displayed a lower PAE (β = −.617) than students in the middle educational track. Among all predictors on the person level, mathematical achievement was more strongly related to PAE (**β** = −.564) than all other person-level characteristics (all βs ≤ |.284|). On the trial level, greater distance from the nearest orientation points (β = .111) and a higher position in the presentation order (β = .101) were associated with a higher PAE. When all predictors were entered as predictors in a multiple regression, mathematical achievement predicted number line estimation PAE over and above the other predictors with β =−.411. Again, as expected, the effect sizes found in the bivariate regressions were higher than the ones found in the multiple regressions due to the multicollinearity of the predictors.

We also examined to what extent the relation between number line estimation PAE and mathematical achievement differed between the quartiles of mathematical achievement. To this end, we computed, separately for each quartile of achievement, a bivariate three-level regression with mathematical achievement as a predictor of number line estimation PAE. The resulting regression coefficients are listed in Table S4 of the supplementary material, where the relations are visualized in Figs. S15 and S16. Due to the restricted variance, the relations were generally weaker in the quartiles than in the overall population. Whole-number estimation and achievement were more closely related in the first achievement quartile (i.e., for low-achieving students) than in the other quartiles. Fraction estimation and achievement were more closely related in the fourth achievement quartile (i.e., for high-achieving students) than in the other quartiles.

Discussion

Research Question 1: Comparison standards

We provided benchmarking data for number line estimation accuracy in the population of Luxemburgish ninth-graders and several relevant subpopulations (students differing in their educational tracks, quartiles of mathematical achievement, gender, or migration background) (Research Question 1). The subpopulations differed substantially in the mean accuracy of their estimates. For example, students on the highest educational track (PAE = 5.55) estimated fractions about four times more accurately than students on the lowest track (PAE = 22.55). Students in the highest mathematical achievement quartile (PAE = 4.56) estimated fractions about four times more accurately than students in the lowest achievement quartile (PAE = 20.64). This shows that subpopulation-specific comparison standards can be useful when interpreting a person’s number line estimation accuracy.

Research Question 2: Number line estimation accuracy as predictor of performance

A meta-analysis had found strong bivariate correlations between number line estimation and mathematical achievement (Schneider et al., 2018). We successfully replicated the finding that number line estimation and achievement are correlated (see Ellis et al., 2021, for another successful replication attempt). However, the standardized regression coefficients from bivariate regressions found in our study were somewhat lower (.225 for whole numbers and .384 for fractions) than the correlations in the meta-analysis (.381 for whole numbers and .529 for fractions in children older than 9 years). There are at least three differences between the meta-analysis and our study that can explain why the effect sizes were smaller in our study. First, the meta-analytic results were upward corrected for measurement unreliability, whereas the current results were not. Previous studies not adjusting their results upward for measurement unreliability reported associations between number line estimation and mathematical achievement similar to the ones reported in the current study (e.g., r = −.34 for whole numbers: Tosto et al., 2017; r = .42 for fractions: Rodrigues et al, 2019). Second, the correlation between number line estimation and achievement increases with the number of number line estimation trials used in the studies (Schneider et al., 2018). The studies in the meta-analysis used 22 trials on average, whereas we used only 10 trials with whole numbers and 9 trials with fractions. Finally, in the meta-analysis, most sample mean ages were younger than 14 years, whereas our participants here had a mean age of 14 years. For whole-number estimation, the meta-analysis had found that the correlation is strongest for 6- to 9-year-old children and is weaker for younger and older children. For fraction estimation, the meta-analysis had found that the correlation is higher for children older than 9 years than for children aged 6 to 9 years. Together, these results suggest that the number line estimation task with whole numbers between 0 and 1000 might be more useful when testing primary school students and that the task with fractions between 0 and 1 might be more useful when testing secondary school students.

Going beyond the meta-analytic findings, we showed that the estimation–achievement relation weakens but does not vanish when it is controlled for a wide range of person characteristics and classroom characteristics. This leaves open the possibility that number line estimation is causally related to mathematical achievement. It remains a task for future studies to test hypotheses about causation in experimental designs.

Research Question 3: Sources of individual differences in number line estimation accuracy

We examined which predictors on the classroom level, person level, and trial level explained variance proportions of number line estimation accuracy. A high level of covariance between number line estimation and achievement would indicate a high convergent validity of the number line estimation task as a measure of mathematical achievement. Low levels of covariance between number line estimation and measures other than achievement would indicate a high discriminant validity of the number line estimation task as a measure of mathematical achievement. We found that mathematical achievement explained proportions of the between-person variance in number line estimation of 14.2% for whole numbers and 31.8% for fractions. Each other person characteristic explained less than 3% of the variance for whole numbers and less than 9% of the variance for fractions. Number line estimation was strongly related to the education track. This was not surprising because students with higher aptitude attend higher tracks and mathematical instruction is more demanding on the higher tracks. Overall, the number line estimation task had satisfactory convergent validity and excellent discriminant validity in that it reflected mathematical achievement to a much higher degree than all other person characteristics, that is, mathematical self-concept and anxiety, test motivation, conscientiousness, gender, SES, and migration status. Trial characteristics explained a proportion of 10.0% of the variance on the trial level. This indicates that the characteristics of the presented numbers need to be considered when using the number line estimation task as a component of standardized assessments. We return to this point below.

The number line estimation task as a diagnostic tool

What are the implications of this study for the use of number line estimation tasks as diagnostic tools for mathematical achievement? Overall, our findings support the validity of the number line estimation task as part of mathematical achievement tests or as a short assessment to approximate mathematical achievement in situations where a full test cannot be used. Obviously, short assessments have less desirable psychometric qualities than full tests. This can be seen here.

Number line estimation with whole numbers explained 5.1% of the variance of mathematical achievement, and number line estimation with fractions explained 19.3% of the variance of mathematical achievement. By the standards set by Cohen (1992), these effect sizes are small to medium for whole numbers and medium to large for fractions. So, as expected, the number line estimation task cannot fully replace a 40-min, standardized, curricularly valid Rasch scaled achievement test. This is further supported by the classification accuracy; when we used the number line estimation accuracy quartiles to predict the mathematical achievement quartiles, the classification accuracy was 38.9% for whole numbers and 43.5% for fractions. These values are far from being perfect. However, they are still high when considering that we used only 10 trials of number line estimation with whole numbers and only 9 trials of number line estimation with fractions. Most previous studies used more number line estimation trials, which increases the correlation between number line estimation and achievement (Schneider et al., 2018). Thus, we likely would have found higher correlations in our study if we had included more trials.

The correlation between number line estimation and achievement found in our study is substantially strong in comparison with the effect sizes found for other variables frequently used as predictors or covariates of mathematical achievement. Number line estimation explained variance proportion of achievement of 5.1% for whole numbers and 19.3% for fractions in our study. By comparison, previous meta-analyses found that nonsymbolic magnitude comparison explains a variance proportion of mathematical achievement of 5.8% (Schneider et al., 2017), symbolic magnitude comparison explains 9.1% (Schneider et al., 2017), working memory explains 12.3% (Peng et al., 2016), spatial ability explains 7.3% (Xie et al., 2020), and fluid intelligence 16.8% (Peng et al., 2019). For arithmetic, we did not find a meta-analysis. Nunes et al. (2011) reported that arithmetic explained about 5% to 8% of the variance in mathematical achievement over and above intelligence and working memory in a large sample of British school students. Torbeyns et al. (2015) found that number line estimation accuracy with fractions predicted mathematical achievement better than fraction arithmetic did. This finding was consistent over small samples of Belgian, Chinese, and U.S. sixth- and eighth-graders. Thus, 9 trials of number line estimation with fractions predict mathematical achievement better than many other commonly used predictors of achievement and about equally well as standardized intelligence tests, which comprise many more tasks and take much more time than our number line task. Seen this way, the number line task is far from being perfect and yet, at the same time, is highly useful for diagnostic purposes.

The measurement of mathematical achievement is a central component of the diagnostic process for dyscalculia or mathematical giftedness. We had no additional measures of dyscalculia or giftedness, so we could not cross-validate the findings obtained with our mathematical achievement test. However, at least for dyscalculia, there are studies that investigated smaller samples in greater detail and found that dyscalculic learners have lower accuracy and use less effective solution strategies in number line estimation (e.g., Lafay et al., 2016; van't Noordende et al., 2016). These findings cross-validate our own findings.

Which number line task version should be used for ninth-graders? Overall, mathematical achievement was more strongly related to the PAE in the fractions number line estimation task ($r = -.533$) than in the whole-numbers number line estimation task ($r = -.359$). This result mirrors findings from the previous meta-analysis (Schneider et al., 2018). This stronger association between fractions number line estimation accuracy and mathematical achievement might be explained through overlapping competencies that are required for the fractions number line estimation task and mathematical achievement tests (Link et al., 2014). A higher sensitivity of the fractions number line estimation task for capturing differences in mathematical achievement is also indicated by the pattern of differences between different subpopulations; differences between students from different educational tracks and quartiles of mathematical achievement were more strongly pronounced for the fractions number line task than for the whole-numbers number line task. Thus, the fractions number line estimation task may be useful for identifying both dyscalculic and mathematically gifted students and for assessing mathematical achievement in students with average levels of mathematical achievement.

What are the implications of the influence of trial-level characteristics on number line estimation accuracy for designing and comparing number line estimation tasks? The analysis of variance components revealed that a substantial proportion of variance in number line estimation occurs on the trial level (67.1% for whole numbers, 70.9% for fractions) compared with the classroom level and person level, which played a smaller role. This variance could systematically be explained by trial-level characteristics such as the distance from the nearest orientation point across the two versions of the number line estimation task. Thus, number line estimation accuracy depends on characteristics of the numbers that are included in the respective number line estimation task. This finding has two major implications. First, researchers and practitioners may systematically vary the difficulty of the applied number line estimation tasks by altering trial-level characteristics (e.g., by including numbers with a larger distance to orientation points). Second, this finding shows that estimation accuracy cannot be compared across studies using different stimuli in the number line estimation task (see Lai et al., 2018, for further evidence). Thus, to keep the results comparable across studies, it might be useful to create standardized sets of stimuli (numbers) for assessing number line estimation accuracy. To avoid ceiling or floor effects, different sets of numbers need to be created for different populations (e.g., primary school students, secondary school students), and the number type, number range, and other trial characteristics need to be carefully chosen to be adequate for this population.

Relations of number line estimation with further person-level variables

Although number line estimation accuracy was most strongly related to mathematical achievement on the person level, some noteworthy relations with motivational and demographic characteristics emerged. This study provided the first evidence for positive relations of mathematical self-concept and test motivation with number line estimation accuracy. They show that number line estimation accuracy is not purely determined by cognitive variables. We also found that gender and migration status were related to number line estimation accuracy. Previous research found that a male advantage in mathematics could be partly explained by more positive mathematical attitudes and self-efficacy (e.g., Else-Quest et al., 2010). However, our results showed that differences in motivational variables do not fully account for gender differences in number line estimation accuracy. Further research on the network of interrelations between cognitive mathematical processes, motivational processes, and language in number line estimation is needed.

Limitations and future directions

There are several limitations to this study. First, we could not control for cognitive correlates of number line estimation accuracy and mathematical achievement in our analyses. Thus, we cannot rule out the possibility that domain-general cognitive abilities account for the relation of number line estimation accuracy and mathematical achievement in our sample. However, previous studies have demonstrated unique relations of number line estimation accuracy with mathematical competencies beyond variables such as intelligence (e.g., Siegler & Booth, 2004) and working memory capacity (e.g., Zhu et al., 2017). In addition, even if the number line estimation accuracy–mathematical achievement relation were accounted for by domain-general cognitive abilities, the number line estimation task would still be a useful screener for mathematical achievement due to its briefness (~4 min total in our study) compared with measures of domain-general cognitive abilities and other measures of mathematical achievement.

Second, we used a cross-sectional design, so we cannot draw any conclusions regarding causality. Longitudinal research and experimental research are needed to investigate the developmental interplay of number line representation, mathematical achievement, and motivational and demographic variables. Such studies could also test the influence of classroom-level characteristics on the codevelopment of number line representation and mathematical achievement.

Third, we did not systematically assess classroom-level variables (e.g., teacher competence, characteristics of classroom instruction). Thus, it is unclear which characteristics of the classroom or the instruction are responsible for the large between-classroom variance. This remains an open question for future research.

Fourth, our sample was homogeneous in that all students were from the same school year and country. Future research is needed to test the generalizability of our findings to countries differing from Luxembourg in school forms or student demographics. There are reasons to expect high generalizability. In *PISA 2015,* students in Luxembourg had a mathematical literacy score of 486, which is very similar to the OECD (Organization for Economic Cooperation and Development) overall mean of 490 (Boehm et al., 2016). Thus, at least in terms of achievement, the Luxembourgish students are representative for students in the OECD. Our findings from Luxembourg are also well in line with the findings of a meta-analysis aggregating data from 11 countries in North America, Europe, and Asia (Schneider et al., 2018).

Conclusion

This study showed that subpopulations differ in their number line estimation accuracy, so subpopulation-specific norms might be helpful when using the number line estimation task as a diagnostic tool for mathematical achievement. The number line estimation task has an acceptable convergent validity in that it predicts mathematical achievement better than or as well as nonsymbolic magnitude comparison, symbolic magnitude comparison, spatial ability, working memory, and fluid intelligence. This is remarkable because number line estimation requires fewer trials and less time than most of the other measures. The number line estimation task had an excellent discriminant validity in that it reflected mathematical achievement to a much greater extent than it reflected any other person characteristic investigated in our study. Thus, the number line estimation task can be used as a brief diagnostic tool for predicting and monitoring mathematical achievement. In secondary school students and high-achieving students, the number line estimation task with fractions has a higher sensitivity for individual differences than the task with whole numbers.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2022.105521.

References

Agirdag, O., van Houtte, M., & van Avermaet, P. (2012). Why does the ethnic and socioeconomic composition of schools influence math achievement? The role of sense of futility and futility culture. *European Sociological Review, 28*(3), 366–378. https://doi.org/10.1093/esr/jcq070.

Alcock, L., Attridge, N., Kenny, S., & Inglis, M. (2014). Achievement and behavior in undergraduate mathematics: Personality is a better predictor than gender. *Research in Mathematics Education, 16*(1), 1–17. https://doi.org/10.1080/14794802.2013.874094.

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science, 17*(5), 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*(1), 125–135. https://doi.org/10.1111/j.1467-7687.2010.00962.x.

Boehm, B., Fischbach, A., Keller, U., Lorphelin, D., Martin, R., & Ugen, S. (2016). *PISA 2015: Nationaler Bericht*. Ministerium für Bildung, Kinder und Jugend, Service de Coordination de la Recherche et de l'Innovation pédagogiques et technologiques (SCRIPT), Luxembourg. https://www.script.lu/de/publications?field_categorie_target_id=9.

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.

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology, 102*, 964–981. https://doi.org/10.1037/a0019644.

Cohen, D. J., & Ray, A. (2020). Experimental bias in number-line tasks and how to avoid them: Comment on Kim and Opfer (2017) and the introduction of the Cohen Ray number-line task. *Developmental Psychology, 56*, 846–852. https://doi.org/10.1037/dev0000761.

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

Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*(6), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I.

Cowger, C. D. (1984). Statistical significance tests: Scientific ritualism or scientific method? *Social Service Review, 58*, 358–372.

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.

Ellis, A., Susperreguy, M. I., Purpura, D. J., & Davis-Kean, P. E. (2021). Conceptual replication and extension of the relation between the number line estimation task and mathematical competence across seven studies. *Journal of Numerical Cognition, 7*(3), 435–452. https://doi.org/10.5964/jnc.7033.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. https://doi.org/10.1037/a0018053.

Fischbach, A., Ugen, S., & Martin, R. (2014). ÉpStan technical report. *University of Luxembourg* https://orbilu.uni.lu/handle/10993/15802.

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socioeconomic index of occupational status. *Social Science Research, 21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology, 47*(6), 1539–1552. https://doi.org/10.1037/a0025510.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*(3), 277–299. https://doi.org/10.1080/87565640801982361.

Gersten, R., Clarke, B. S., Haymond, K., & Jordan, N. C. (2011). *Screening for mathematics difficulties in K–3 students* (2nd ed.). Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Sage. https://doi.org/10.4135/9781412986311.n21.

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational–affective constructs with three-item and single-item measures. *Contemporary Educational Psychology, 39*(3), 188–205. https://doi.org/10.1016/j.cedpsych.2014.04.002.

Gouet, C., Carvajal, S., Halberda, J., & Peña, M. (2020). Training nonsymbolic proportional reasoning in children and its effects on their symbolic math abilities. *Cognition, 197*, 104154. https://doi.org/10.1016/j.cognition.2019.104154.

Gravetter, F., & Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Boston: Cengage Learning.

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*(5), 1229–1241. https://doi.org/10.1037/a0027433.

Hart, S. A., Logan, J. A., Thompson, L., Kovas, Y., McLoughlin, G., & Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology, 108*(2), 181–193.

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *Journal of Cross-Cultural Psychology, 42*(4), 598–613. https://doi.org/10.1177/0022022111406026.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834–850. https://doi.org/10.1111/1467-8624.00571.

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*(1), 45–58. https://doi.org/10.1016/j.jecp.2013.02.001.

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews, 15*(1), 60–68. https://doi.org/10.1002/ddrr.46.

Keller, U., Strobel, A., Wollschläger, R., Greiff, S., Martin, R., Vainikainen, M. P., & Preckel, F. (2016). A need for cognition scale for children and adolescents. *European Journal of Psychological Assessment, 35*(1), 137–149. https://doi.org/10.1027/1015-5759/a000370.

Kodosh, R. C., Tzelgov, J., & Henik, A. (2008). A synesthetic walk on the mental number line: The size effect. *Cognition, 106*(1), 548–557. https://doi.org/10.1016/j.cognition.2006.12.007.

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review, 25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001.

Kucian, K., McCaskey, U., Tuura, R. O. G., & von Aster, M. (2018). Neurostructural correlate of math anxiety in the brain of children Article 273. *Translational Psychiatry, 8*. https://doi.org/10.1038/s41398-018-0320-6.

Lafay, A., St-Pierre, M.-C., & Macoir, J. (2016). The mental number line in dyscalculia: Impaired number sense or access from symbolic numbers? *Journal of Learning Disabilities, 50*, 672–683. https://doi.org/10.1177/0022219416640783.

Lai, M., Zax, A., & Barth, H. (2018). Digit identity influences numerical estimation in children and adults Article e12657. *Developmental Science, 21*. https://doi.org/10.1111/desc.12657.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition, 93*(2), 99–125. https://doi.org/10.1016/j.cognition.2003.11.004.

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

LUCET (Luxembourg Center for Educational Training) (n.d.). *ÉpStan-Secondaire (5e): Beispielaufgaben aus dem Mathematiktest.* Université du Luxembourg. https://epstan.lu/downloadbereich-9.

Lüftenegger, M., Kollmayer, M., Bergsmann, E., Jöstl, G., Spiel, C., & Schober, B. (2015). Mathematically gifted students and high achievement: The role of motivation and classroom structure. *High Ability Studies, 26*(2), 227–243. https://doi.org/10.1080/13598139.2015.1095075.

Marsh, H. W., & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement, 21*(2), 153–174. https://doi.org/10.1111/j.1745-3984.1984.tb00227.x.

Martin, R., Ugen, S. & Fischbach, A. (2015). *Épreuves Standardisées. Nationaler Bericht 2011 bis 2013.* Universität Luxemburg, LUCET. Retrieved December 2, 2020, from https://orbilu.uni.lu/handle/10993/13647.

MENFP (Ministère de l'Education nationale et de la Formation professionnelle) (2008). *Mathématiques: Division inférieure de l'enseignement secondaire technique: Compétences disciplinaires attendues à la fin de la classe de 6eet à la fin de la classe de 4 e avec compétences élargies pour l'accès aux sections B, C et D*. [Mathematics: Lower grades of secondary education: Disciplinary competences expected at the end of 6th and 4th grade with extended competences for access to sections B, C and D]. MENFP, Luxembourg.

Muthén, B. O., & Muthén, L. K. ((1998–2017).). *Mplus user's guide* (8th ed.). Los Angeles: Muthén & Muthén.

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research, 89*(3), 459–496. https://doi.org/10.3102/0034654319843494.

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2011). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology, 82*(1), 136–156. https://doi.org/10.1111/j.2044-8279.2011.02033.x.

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology, 34*(1), 29–48. https://doi.org/10.1080/01443410.2013.797339.

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*(4), 455–473. https://doi.org/10.1037/edu0000079.

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin, 145*(2), 189–236. https://doi.org/10.1037/bul0000182.

Penk, C., & Richter, D. (2017). Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educational Assessment, Evaluation and Accountability, 29*(1), 55–79. https://doi.org/10.1007/s11092-016-9248-7.

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction, 7*(1), 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*(2), 375–394. https://doi.org/10.1111/j.1467-8624.2007.01131.x.

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

Rodrigues, J., Jordan, N. C., & Hansen, N. (2019). Identifying fraction measures as screeners of mathematics risk status. *Journal of Learning Disabilities, 52*(6), 480–497. https://doi.org/10.1177/0022219419879684.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis Article e12372. *Developmental Science, 20*(3). https://doi.org/10.1111/desc.12372.

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology, 101*(2), 359–372. https://doi.org/10.1037/a0013840.

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., de Smedt, B., Verschaffel, L., Jacobs, A. M., & Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development, 23*(3), 409–422. https://doi.org/10.1016/j.cogdev.2008.07.002.

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development, 89*(5), 1467–1484. https://doi.org/10.1111/cdev.13068.

Sheffield, L. J. (1994). *The development of gifted and talented mathematics students and the national council of teachers of mathematics standards.* Connecticut, OH: The National Research on the Gifted and Talented.

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., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations. *Psychological Science, 14*(3), 237–243. https://doi.org/10.1111/1467-9280.02438.

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273–296. https://doi.org/10.1016/j.cogpsych.2011.03.001.

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*(3), 557–563. https://doi.org/10.3758/s13423-011-0081-1.

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.

Tosto, M. G., Garon-Carrier, G., Gross, S., Petrill, S. A., Malykh, S., Malki, K., Hart, S. A., Thompson, L., Karadaghi, R. L., Yakovlev, N., Tikhomirova, T., Opfer, J. E., Mazzocco, M. M. M., Dionne, G., Brendgen, M., Vitaro, F., Tremblay, R. E., Boivin, M., & Kovas, Y. (2018). The nature of the association between number line and mathematical performance: An international twin study. *British Journal of Educational Psychology, 89*(4), 787–803. https://doi.org/10.1111/bjep.12259.

Tosto, M. G., Petrill, S. A., Malykh, S., Malki, K., Haworth, C. M. A., Mazzocco, M. M. M., Thompson, L., Opfer, J., Bafdanova, Y. K., & Kovas, Y. (2017). Number sense and mathematics: Which, when and how? *Developmental Psychology, 53*(10), 1924–1939. https://doi.org/10.1037/dev0000331.

van't Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., & Kroesbergen, E. H. (2016). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychological Research, 80*, 368–378. https://doi.org/10.1007/s00426-015-0736-z.

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.

Wang, Z., Lukowski, S. L., Hart, S. A., Lyons, I. M., Thompson, L. A., Kovas, Y., Mazzocco, M. M. M., Plomin, R., & Petrill, S. A. (2015). Is math anxiety always bad for math learning? The role of math motivation. *Psychological Science, 26*(12), 1863–1876. https://doi.org/10.1177/0956797615602471.

Xie, F., Zhang, L., Chen, X., & Xin, Z. (2020). Is spatial ability related to mathematical ability: A meta-analysis. *Educational Psychology Review, 32*, 113–155. https://doi.org/10.1007/s10648-019-09496-y.

Yuan, L., Prather, R., Mix, K. S., & Smith, L. B. (2020). Number representations drive number-line estimates. *Child Development, 91*(4), e952–e967. https://doi.org/10.1111/cdev.13333.

Zax, A., Slusser, E., & Barth, H. (2019). Spontaneous partitioning and proportion estimation in children's numerical judgments. *Journal of Experimental Child Psychology, 185*, 71–94. https://doi.org/10.1016/j.jecp.2019.04.004.

Zhu, M., Cai, D., & Leung, A. W. (2017). Number line estimation predicts mathematical skills: Difference in Grades 2 and 4 Article 1576. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01576.

Ziegler, A., & Stoeger, H. (2010). How Fine Motor Skills Influence the Assessment of High Abilities and Underachievement in Math. *Journal for the Education of the Gifted, 34*(2), 195–219.