Journal of Experimental Child Psychology 212 (2021) 105251

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Relations among spatial skills, number line estimation, and exact and approximate calculation in young children

Elizabeth A. Gunderson [a,*], Lindsey Hildebrand [b]

[a] Department of Psychology, Temple University, Philadelphia, PA 19122, USA
[b] Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA 02467, USA

ARTICLE INFO

Article history:
Received 11 September 2020
Revised 25 June 2021

Keywords:
Spatial skills
Numeracy
Calculation
Mental rotation
Proportional Reasoning
Number line

ABSTRACT

Decades of research have established that spatial skills correlate with numerical skills. However, because both spatial and numerical skills are multidimensional, we sought to determine how specific spatial skills relate to specific numeracy skills. We used a cohort-sequential design, assessing a large diverse sample of students (N = 612, initially in pre-kindergarten [pre-K]–3rd grade, 4–9 years of age) at four time points spanning 2 years. We examined how initial levels of five spatial skills (visuospatial working memory [VSWM], mental transformation, mental rotation, proportional reasoning, and analog magnitude system [AMS] acuity) related to initial levels and growth rates in exact and approximate calculation skills, and we further investigated number line estimation as a potential mediator. We found unique patterns of relations between spatial skills and numeracy. Initial levels of mental rotation, proportional reasoning, and AMS acuity related to initial levels of exact calculation skill; initial levels of AMS acuity related to initial levels of approximate calculation; and initial levels of proportional reasoning related to initial levels of number line estimation. VSWM and mental transformation did not relate to numeracy skills after controlling for other spatial skills. Initial levels of number line estimation related to both exact and approximate calculation after controlling for spatial skills. Notably, neither spatial skills nor number line estimation predicted growth in exact or approximate calculation skills. These results indicate that there is specificity in the time-invariant relations between spatial skills and numeracy, and they suggest that researchers and educators should

* Corresponding author.
E-mail address: liz.gunderson@temple.edu (E.A. Gunderson).

https://doi.org/10.1016/j.jecp.2021.105251

treat spatial skills and numeracy as multidimensional constructs with complex and unique interrelations.

Introduction

Mathematics achievement is a critical predictor of academic success and career outcomes, including adulthood income (Ritchie & Bates, 2013; Watts, Duncan, Siegler, & Davis-Kean, 2014). Spatial skills—a suite of abilities that allow individuals to encode, hold in mind, and mentally manipulate spatial information—are a major predictor of achievement, interest, and success in mathematics as early as 4 years of age (Casey, Nuttall, Pezaris, & Benbow, 1995; Kyttälä, Aunio, Lehto, Van Luit, & Hautamäki, 2003; Lachance & Mazzocco, 2006). Importantly, spatial skills can be substantially improved through training at all ages (Uttal et al., 2013). Therefore, enhancing spatial skills may represent a powerful and currently underutilized leverage point for enhancing children's early numeracy skills (Hawes, Moss, Caswell, Naqvi, & MacKinnon, 2017).

However, to leverage spatial skills to create effective interventions, it is critical to first understand *how* and *why* spatial skills are related to specific numerical abilities (Mix & Cheng, 2012; Uttal & Cohen, 2012). Despite more than 50 years of research establishing the relation between spatial skills and later STEM (science, technology, engineering, and mathematics) achievement (Wai, Lubinski, & Benbow, 2009), researchers have only recently begun to delineate the specific mechanisms through which spatial skills relate to symbolic numeracy (Gunderson, Ramirez, Beilock, & Levine, 2012; Mix et al., 2016).

The current study used a large 2-year cohort-sequential study of children in pre-kindergarten (pre-K) to 4th grade (4–10 years of age) to test predictions about *how* specific spatial skills relate to specific aspects of symbolic numeracy (approximate and exact calculation). Furthermore, we examined *why* spatial skills relate to symbolic numeracy by testing one proposed mechanism—that spatial skills improve children's spatial–numeric representation of numbers (e.g., a mental number line)—which in turn improves other numerical skills (Gunderson et al., 2012; LeFevre et al., 2013). With this proposed mechanism in mind, we examined spatial skills that have been theoretically and empirically linked to number line estimation skill: visuospatial working memory (VSWM), mental transformation, mental rotation, proportional reasoning, and analog magnitude system (AMS) acuity.

Exact and approximate symbolic numeracy skills

We define "symbolic numeracy skills" as mathematical concepts and procedures that involve number symbols (i.e., Arabic numerals or number words; e.g., De Smedt, Noël, Gilmore, & Ansari, 2013). In pre-K to 4th grade, these skills typically involve understanding the magnitudes represented by number symbols and manipulating these number symbols using arithmetic operations. A salient cleavage in symbolic numeracy skills is the distinction between approximate and exact symbolic numeracy skills (Caviola, Mammarella, Cornoldi, & Lucangeli, 2012; Spelke & Tsivkin, 2001; Waring & Penner-Wilger, 2017).

Approximate symbolic numeracy skills, also referred to as "estimation" or "number sense," rely heavily on children's knowledge of the magnitudes represented by number symbols. A hallmark of approximate symbolic numeracy skill is that when comparing two numbers, performance is better the larger the ratio between the two numbers to be compared (i.e., ratio dependent; De Smedt et al., 2013). Children can also perform arithmetic operations approximately even before formal instruction (Gilmore, McCarthy, & Spelke, 2007; Xenidou-Dervou, De Smedt, van der Schoot, & van Lieshout, 2013). For example, 5- and 6-year-olds can successfully choose the larger of "21 + 30 versus 34" without being able to say that "21 + 30 = 51" (Gilmore et al., 2007). This approximate symbolic calculation skill is correlated with general math achievement (Xenidou-Dervou et al., 2013) and is predicted by children's earlier spatial skills, a relation that is mediated by number line estimation skills (Gunderson et al., 2012).

In contrast to approximate symbolic numeracy skills, exact numerical skills involve manipulating number symbols to arrive at a precise answer, typically using a combination of retrieval and procedural rules. In the United States, a focus on exact calculation ability is present during the first years of formal schooling, and most standardized math assessments involve exact calculations (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Importantly, both approximate and exact numerical skills are strong predictors of math achievement (e.g., Siegler et al., 2012; Xenidou-Dervou et al., 2013). Yet, these are distinct numerical skills that we expect to have distinct relations to specific spatial skills.

Role of number line estimation in math achievement

Prior research has identified number line estimation as a potential mediator of relations between spatial skills and symbolic numeracy skills (Gunderson et al., 2012). Number line estimation skill—the ability to represent the magnitude of a target number on a line with two labeled end points—is strongly related to other numerical skills, including recalling numbers, fraction knowledge, approximate calculation, and symbolic estimation (for a review, see Siegler & Lortie-Forgues, 2014). Furthermore, a recent meta-analysis found that number line estimation skill was positively related to math achievement during childhood ($r = .443$) (Schneider et al., 2018). Here, we expected number line estimation skill to be more strongly related to approximate symbolic numeracy skill than to exact symbolic numeracy skill because increased precision of the symbolic number sense is particularly important for success on symbolic approximation tasks. Indeed, children's number line knowledge predicts their later approximate symbolic calculation skill (Gunderson et al., 2012), and training number line knowledge leads to improvements in other symbolic approximation skills such as magnitude comparison (Ramani & Siegler, 2008) and approximate arithmetic (Booth & Siegler, 2008).

In contrast to approximate symbolic calculation, exact calculation requires knowledge of specific math facts and procedures. Although children's number magnitude representations may indirectly benefit their exact calculation skills (e.g., by helping them to quickly check whether a calculated solution is reasonable), number magnitude representations alone cannot be used to perform exact calculations. Consistent with this, approximate (but not exact) numerical instruction has been shown to improve children's approximate calculation performance (Obersteiner, Reiss, & Ufer, 2013). Therefore, we expected to find weaker (or no) relations between number line estimation and exact calculation skills given the approximate nature of number line estimation.

Specific spatial skills

Past work comparing spatial and numerical skills suggests that skills in these two domains are distinct but related. For example, researchers have used factor analyses to show that math and spatial skills load onto distinct but related factors during early and middle childhood (Hawes, Moss, Caswell, Seo, & Ansari, 2019; Mix et al., 2016, 2017). Furthermore, work examining relations between multiple spatial skills and math achievement suggests unique relations between specific spatial skills and math outcomes (Geer, Quinn, & Ganley, 2019; Gilligan, Hodgkiss, Thomas, & Farran, 2019). These findings highlight the importance of taking a broad approach to studying the multidimensional interplay between spatial and mathematical development.

Studies examining psychometrically defined "spatial skills," such as mental rotation, have often been conducted in isolation from other spatially relevant skills. The current study considered spatial skills more broadly and included both psychometrically defined skills (mental rotation and mental transformation) and other spatially relevant skills with strong ties to symbolic numeracy, including VSWM, proportional reasoning, and acuity of the AMS. Including a broad set of spatially relevant skills in a single study is critical for a robust understanding of the unique contributions of spatial skills to symbolic numeracy. We chose these spatial skills because of their theoretical and empirical relations to number line estimation and predicted that each skill would have an *indirect* effect on approximate symbolic numeracy via number line estimation skills. In addition, we expected that some of these spatial skills

(VSWM, mental rotation, and mental transformation) would also have *direct* effects on exact symbolic numeracy skills. Here, we explain the theorized mechanisms underlying these predictions.

Mental transformation and mental rotation

Mental rotation is the ability to hold in mind and mentally rotate representations of two- or three-dimensional visual stimuli (e.g., identifying a rotated letter; Shepard & Metzler, 1971). We use the term "mental rotation" to refer strictly to tasks that require children to distinguish between rotated versions of mirror-reversed images; this eliminates the possibility of strategies based on matching specific features of the images. We use the term "mental transformation" to refer to tasks that may involve mentally rotating or mentally translating visual stimuli but can also be solved using perceptual information such as feature matching (e.g., identifying which puzzle piece would fit in a puzzle). Most prior work relating spatial skills to numeracy treats mental transformation and mental rotation as interchangeable. In a recent exception, Frick (2019) found that kindergartners' mental transformation and mental rotation skills were not correlated after controlling for age, verbal IQ, and socioeconomic status (SES). Furthermore, these skills loaded on different spatial factors with distinct relations to numeracy, suggesting that these are in fact distinct skills. We included measures of both skills to test whether one is a more robust predictor of children's symbolic numeracy skills.

We expected mental transformation and mental rotation to contribute to approximate symbolic numeracy indirectly via number line estimation skills. Theoretically, visual transformation strategies such as "zooming" between more and less familiar scales, a strategy children use on nonsymbolic scaling tasks (Möhring, Newcombe, & Frick, 2014), may enhance children's ability to visualize magnitudes on the number line task. Empirically, spatial skills requiring mental transformation and mental rotation predict later number line estimation skill during early childhood (Gunderson et al., 2012; LeFevre et al., 2013). As noted above, we expected number line estimation skills in turn to predict children's approximate (but not exact) symbolic calculation skills. We did not have a priori predictions regarding whether mental rotation alone, mental transformation alone, or both would drive these predicted relations, and we included both types of measures to explore these relations.

In addition, we expected mental transformation and mental rotation to relate directly to exact symbolic numeracy skills by supporting the mental models through which children solve novel math problems (referred to as the "spatial modeling hypothesis"; Hawes, Moss, et al., 2019; Mix, 2019). Spatial mental models can ground novel problems in a visuospatial representation that is easier for learners to reason about and manipulate than the symbolic equivalent and, as such, can support learning of new mathematical concepts at all ages (Uttal & Cohen, 2012). For example, when first learning addition, children may create a mental model of three objects coming together with two objects to solve the problem "3 + 2." At an older age, constructing a schematic spatial model via sketching was beneficial for 6th-graders solving complex word problems (Hegarty & Kozhevnikov, 1999). Therefore, we expected mental transformation and mental rotation skills to directly relate to exact calculation skills. Consistent with this prediction, elementary school students' mental rotation skill correlates with their exact calculation skills (Georges, Cornu, & Schiltz, 2019; Skagerlund & Träff, 2016), and mental transformation training can lead to improvements in arithmetic calculation (Cheng & Mix, 2014; Mix, Levine, Cheng, Stockton, & Bower, 2021).

Visuospatial working memory

VSWM is a cognitive system that stores and processes information in terms of visual and spatial features (Baddeley & Hitch, 1974). VSWM robustly predicts numeracy skills, including counting and arithmetic, in pre-K to 3rd grade (4–9 years of age; Kyttälä et al., 2003; Meyer, Salimpoor, Wu, Geary, & Menon, 2010). Like mental rotation and transformation, we expected VSWM to affect symbolic numeracy skills both indirectly (via number line estimation) and directly. We suggest that VSWM may aid number line estimation by helping children to hold in mind locations associated with specific numbers. Consistent with this, 8- to 10-year-olds' VSWM, mental rotation, disembedding, and visuomotor integration all correlated with 0–1000 number line estimation (Simms, Clayton, Cragg, Gilmore, & Johnson, 2016). However, to our knowledge, no study has established whether VSWM is a unique predictor of number line estimation over and above other spatial skills. We predicted that VSWM would relate to number line estimation skill, which in turn would predict approximate (but not exact) symbolic calculation skills.

Children may use VSWM to hold in mind and manipulate relevant spatial information when grounding new concepts with spatial mental models. Indeed, ample empirical evidence indicates that VSWM is a strong predictor of overall math achievement (Raghubar, Barnes, & Hecht, 2010) and a specific predictor of exact calculation skills during childhood (e.g., Foley, Vasilyeva, & Laski, 2017). For example, 6- to 9-year-olds rely on VSWM as a mental sketchpad while performing simple mental arithmetic problems (McKenzie, Bull, & Gray, 2003). Furthermore, VSWM can improve children's ability to interpret and mentally manipulate symbols, including Arabic numerals (Mix, 2019). Indeed, recent evidence provides causal support for improvements in 1st- and 6th-graders' arithmetic performance following VSWM training (Mix et al., 2021). Thus, we also expected VSWM to directly relate to exact calculation skills.

Proportional reasoning

Proportional reasoning requires reasoning about visual relations between nonsymbolic spatial extents (e.g., line lengths; Möhring, Newcombe, & Frick, 2015). Early studies by Piaget and Inhelder (1975) suggested that proportional reasoning emerges at around 11 years of age, but more recent research suggests that proportional reasoning may emerge in 4- to 6-year-olds if the presentation of the proportion judgments are continuous amounts (e.g., Boyer, Levine, & Huttenlocher, 2008). Basic proportional reasoning is evident in 3- and 4-year-olds when asked to produce equivalent amounts (e.g., half a pizza equals half a chocolate bar; Singer-Freeman & Goswami, 2001). Here, we hypothesized that proportional reasoning may be one of the most important skills underlying number line estimation skill and in turn would affect approximate symbolic calculation skills.

There are strong theoretical and empirical reasons to believe that number line estimation relies on proportional reasoning skill. Theoretically, proportional reasoning may provide a spatial strategy for reasoning about symbolic numerical magnitudes. Accurate proportional reasoning requires the ability to compare and map relations between continuous spatial magnitudes. Like proportional reasoning, number line estimation requires considering the relations between a part (i.e., the magnitude of the number to be estimated) and a whole (i.e., the total numerical magnitude of the number line as defined by its end points). This numerical part–whole relation must then be mapped onto an equivalent spatial part–whole relation with respect to the linear spatial extent of the number line. Thus, the ability to reason about relations between spatial magnitudes (indexed by nonsymbolic proportional reasoning) should be beneficial for mathematics, especially for tasks like number line estimation that require the mapping of discrete quantities onto continuous extents (Rouder & Geary, 2014). It is worth noting that we do not see the existence of a proportional strategy on number line estimation as being in conflict with a zooming strategy (discussed previously). Rather, consistent with overlapping waves theory (Chen & Siegler, 2000), we expect that there are multiple number line estimation strategies available to children, and the selection of a particular spatial strategy may vary as a function of individual differences, item types, salience, and other factors.

Empirically, proportion judgments in nonsymbolic tasks show a specific pattern of bias, where large proportions are underestimated and smaller proportions are overestimated (Hollands & Dyre, 2000), and mature performance on number line estimation tasks follows the same pattern of response bias (Slusser, Santiago, & Barth, 2013). In one prior study, both number line estimation and a proportion matching task loaded onto a mathematics factor but not a spatial factor, suggesting that there may be similar magnitude mechanisms that drive a relation between the two skills (Mix et al., 2017). Here, we predicted that proportional reasoning would relate to number line estimation over and above other key spatial skills (mental rotation, mental transformation, VSWM, and AMS acuity).

Analog magnitude system acuity

The AMS is a quick, imprecise system of magnitude estimation (Dehaene, 2003) and is typically assessed by asking children to select the more numerous of two dot arrays. The AMS is evolutionarily old (Nieder, Freedman, & Miller, 2002), is present during infancy (Xu & Spelke, 2000), and improves substantially in acuity during early childhood (Halberda & Feigenson, 2008). Although often described as a numerical skill (and indeed the AMS is often referred to as the "approximate number system" or ANS), AMS performance has a clear spatial component, such that children perform worse when spatial extent contrasts with numerical extent and perform better when spatial and numerical extents coincide (Cantrell & Smith, 2013).

AMS acuity correlates with school math achievement in children as young as 3 years (Bonny & Lourenco, 2013; Libertus, Feigenson, & Halberda, 2011). Individual differences in AMS acuity at 6 months of age predict later math test scores in pre-K (4–5 years of age), which strongly suggests that this capacity—present even before children learn to speak—forms a foundation for symbolic numeracy (Starr, Libertus, & Brannon, 2013).

Much debate has centered on whether the nonsymbolic AMS or symbolic number sense (e.g., speed at comparing Arabic numerals) underlies formal math achievement (De Smedt et al., 2013). This debate largely stems from inconsistent findings on the relation between the AMS and math achievement. These discrepancies may arise from factors such as variations across measures, the type of formal math achievement assessed, and developmental patterns (e.g., De Smedt et al., 2013; Inglis, Attridge, Batchelor, & Gilmore, 2011; Schneider et al., 2017). Notably, relations between the AMS and math achievement are most consistent for math tasks that benefit from the ability to quickly choose and flexibly switch between mental calculation strategies, especially early in childhood (e.g., Linsen, Verschaffel, Reynvoet, & De Smedt, 2014; Nys & Content, 2010). Here, we propose that the precision of the AMS facilitates mapping between spatial magnitudes and symbolic numbers (i.e., number line estimation). Number line estimation requires precision of symbolic numerical representations, which in turn relies on a mapping between symbolic numerals and the nonsymbolic AMS (Siegler & Opfer, 2003). Similarly, approximate symbolic calculation skill, which shows the signature ratio dependency of the AMS (Gilmore et al., 2007), is thought to rely on both the precision of the nonsymbolic AMS and the precision of the nonsymbolic-to-symbolic mappings indexed by number line estimation (e.g., Gilmore, Attridge, De Smedt, & Inglis, 2014; Gilmore et al., 2007; Waring & Penner-Wilger, 2017). Therefore, we predicted that children with more precise nonsymbolic AMS acuity would have better performance on number line estimation and approximate symbolic calculation tasks.

The current study

The current study examined the relations between multiple measures of spatial skills, number line knowledge, and numeracy skills, with the goals of examining *how* (through what specific links) and *why* (through what mechanism) spatial skills predict numerical development. We used a cohort-sequential design in which we recruited a large, racially and socioeconomically diverse cross-sectional sample of children in pre-K to 3rd grade (4–9 years of age), who we assessed twice per year for 2 years. We chose these grade levels because this represents an important time for the development of both spatial and numerical cognition (Uttal & Cohen, 2012). We identified five spatial skills that we believed would predict symbolic numeracy via number line estimation: VSWM, mental transformation, mental rotation, proportional reasoning, and AMS acuity. We examined how these skills related to approximate symbolic calculation as well as exact calculation. We also investigated number line estimation as a potential mediator of the relation of spatial skills to approximate symbolic calculation. We predicted that number line estimation would relate more strongly to approximate calculation than to exact calculation because both number line estimation and approximate calculation are theorized to rely on the precision of symbolic magnitude knowledge (e.g., Gunderson et al., 2012). Nevertheless, we also tested the relation of number line estimation to exact calculation, based on prior empirical findings supporting such a relation (Schneider et al., 2018).

This study presents a unique contribution in its comprehensive examination of a theoretical model (Fig. 1) delineating specific relations between spatial and numerical skills in children using a large-scale longitudinal design. We included several covariates to control for other possible influences on numerical development: reading achievement, verbal IQ, child gender, and parents' SES. We paid special attention to the covariate of child gender given the large literature on the role of gender in math

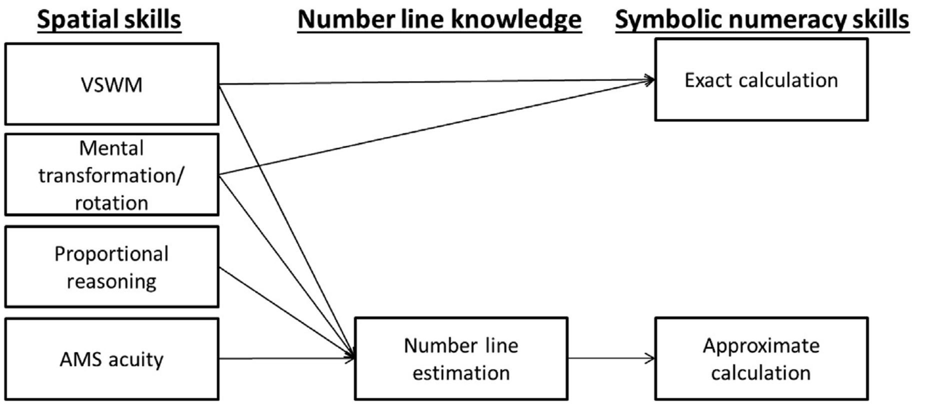

Fig. 1. Theoretical model of relations between spatial and numerical skills. This model summarizes hypothesized links among spatial skills, number line estimation, and symbolic numeracy skills. VSWM, visuospatial working memory; AMS, analog magnitude system.

and spatial development. Specifically, there is a divergence in evidence for gender differences between math and spatial domains. Evidence for gender differences in math skills is minimal (e.g., Bakker, Torbeyns, Wijns, Verschaffel, & De Smedt, 2019; Kersey, Braham, Csumitta, Libertus, & Cantlon, 2018; Spelke, 2005), with some suggesting that findings of gender differences should be considered exceptions rather than the rule (Hutchison, Lyons, & Ansari, 2019). Conversely, there is evidence of a male advantage for spatial skills, especially mental rotation (for a review, see Lauer, Yhang, & Lourenco, 2019). Despite these differences, girls benefit as much as boys from spatial training in improving their spatial and math skills (Uttal et al., 2013), and some evidence suggests that spatial training may even narrow gender gaps (Miller & Halpern, 2013). As such, it is critical to examine the role of gender in studying relations between spatial and numerical skills, especially during this important period of development.

We summarize our theoretical model in Fig. 1. We hypothesized that each spatial skill would uniquely relate to number line estimation. We further expected that number line estimation would relate to approximate symbolic calculation skill and would mediate relations between spatial skills and approximate symbolic calculation but not exact calculation. We predicted that VSWM and mental transformation and/or mental rotation would also relate directly to exact calculation.

We used hierarchical linear models, with time nested within person, to examine whether initial levels of these five spatial skills predicted initial levels and growth rates (i.e., intercepts and slopes) in approximate calculation and exact calculation skills. Finding that *specific* spatial skills related to the initial levels of *specific* numeracy skills would provide support for the hypothesis of multiple distinct links between space and numeracy and would establish the importance of considering both space and numeracy as multidimensional constructs. Finding that specific initial spatial skills predicted *growth* in specific numeracy skills would establish temporal precedence, consistent with the theory that spatial skills lead to enhancements in symbolic numeracy skills.

Although a number of prior studies have reported that earlier spatial skills predict later numerical skills (e.g., Frick, 2019; Gunderson et al., 2012; LeFevre et al., 2013; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017), others have argued that these longitudinal relations do not reflect true temporal precedence and instead are artifacts of relations between spatial skills and numeracy that are stable over time (e.g., Bailey, 2017). Indeed, most studies showing relations between earlier spatial skills and later numerical skills have used linear regression or path analyses, which do not separate stable time-invariant relations from time-varying predictive relations. Our modeling approach separates these relations by treating both the intercept (initial level of numerical skill) and slope (growth rate of numerical skill) as random effects that may vary across individuals. Prior studies using similar statistical approaches, such as by incorporating random intercepts or latent trait components, have found robust time-invariant relations between spatial skills and numeracy and either less robust relations

(Geer et al., 2019) or no evidence for relations that vary over time (Bailey, 2017). Our analytic approach allows us to determine which of two competing hypotheses best characterizes spatial–numeric relations. If spatial–numeric relations are stable over time, as some have argued, this would implicate correlations between stable traits or stable environments—what we call the *time-invariant hypothesis.* Alternatively, spatial–numeric relations may be better characterized by a *spatial–numeric growth hypothesis,* such that early spatial skills not only relate to initial numerical skills but also predict growth rates in numeracy over time. Based on the spatial–numeric growth account, we hypothesized that initial spatial skills would predict both initial levels *and* growth rates in numerical skills, consistent with a causal relation between these skills.

Method

Participants

We recruited participants in fall 2015 (Year 1) from 61 pre-K to 3rd-grade classrooms in 11 schools ($N = 564$). In participating schools, students typically enter pre-K at 4 years of age and enter 3rd grade at 8 years of age (at Time 1 [T1] Session 1, $M_{age} = 6.74$ years, $SD = 1.37$, $n_{age} = 531$; see Appendix Table C1 for mean ages by grade). Students were included in the study if their parents provided informed consent and the children assented to participate. We assessed participants twice per year for 2 consecutive school years (T1 to Time 4 [T4]). We recruited additional participants within these schools between T1 and Time 2 (T2) ($n = 23$) and between T2 and Time 3 (T3) ($n = 40$) and did not contact students who left their school during the study (see Appendix A for an enrollment flow chart). Of the 627 consented participants, we excluded children who did not complete at least one cognitive task ($n = 10$; see Appendix A for details), whose grade level was not recorded ($n = 2$), or who repeated a grade during the study ($n = 3$). Therefore, the analytic sample included 612 children. Based on this sample size, to achieve 80% power with $\alpha = .05$, the minimum detectable effect size in multiple linear regression is $f^2 = .013$, a small effect size (Cohen, 1988).

Students in the analytic sample ($N = 612$) were 338 girls, 273 boys, and 1 child whose gender was not recorded. Parents self-reported family demographics. Children were 51.4% Black/African American, 20.1% White, 11.4% Hispanic, 4.2% Asian/Asian American, 0.4% American Indian/Alaskan Native, 0.4% Native Hawaiian/other Pacific Islander, and 12.1% multiple races/ethnicities ($n_{race/ethnicity} = 543$). Parents' highest level of education averaged 14.67 years, where 14 years is an associate's degree ($SD = 2.38$, $n_{education} = 544$) and levels ranged from less than high school to a graduate degree. Family annual income ranged from less than \$15,000 to more than \$100,000 (M = \$48,524, SD = \$31,633, $n_{income} = 520$). We created a factor score for family SES by entering parents' education and family income into a principal components analysis. The resulting SES factor explained 81% of the variance in parents' education and family income.

Materials and procedure

Students completed assessments of spatial skills, numerical skills, and control measures at each of four time points (see Appendix B for additional task details, including sample trials) approximately 5–7 months apart (T1 to T2: $M = 4.74$ months, $SD = 0.69$; T2 to T3: $M = 7.16$ months, $SD = 1.91$; T3 to T4: $M = 4.82$ months, $SD = 0.32$). At each time point, students completed two 30- to 45-min sessions one on one with a trained experimenter (mean time between sessions = 6.55 days, $SD = 8.56$). At each time point, students completed assessments in one of four pseudorandom orders.[1] All measures were

[1] The measures reported here represent the focal measures that were collected for all grade levels at all time points of this study. The larger longitudinal study included a number of other measures that varied by grade level and/or time point, including measures of whole-number comparison, fraction comparison, 0–10 number line estimation, fraction number line estimation, decimal comparison, linear measurement, counting, cardinality, multiple-choice proportional reasoning, inhibitory control, theories of intelligence, gender stereotypes, academic anxieties, academic self-concepts, academic motivation, and strategy self-reports for spatial and numeric tasks. There is one other published study using this dataset, focusing on the relations between inhibitory control and linear measurement (Ren, Lin, & Gunderson, 2019).

assessed at all four time points with the exception of verbal IQ (assessed only at T1). On tasks with practice trials, participants proceeded to the test trials regardless of their performance on the practice items. No time limits were imposed on any task. See Table 1 for test–retest reliability for each task.

Visuospatial working memory

All participants completed the Dot Matrix subtest of the computerized Automated Work Memory Assessment (AWMA; Alloway, 2007). On each trial, a red dot appeared briefly in a sequence of locations on a 4 × 4 grid, and children were asked to point to the dots' locations in order. This task contained 3 practice trials with feedback. Children received 6 test trials at each span length (number of dots per trial), starting with one dot. The task ended after children gave three wrong answers at the same span length. Each child received a standardized score, normed for 5–69 years of age, computed by the AWMA software. The published test–retest reliability for this task is r = .85 (Alloway, 2007).

Mental transformation

Because there was no measure of mental transformation appropriate for the full age range, we gave a different age-appropriate measure of two-dimensional mental transformation to pre-K and kindergarten children (Children's Mental Transformation Task) than to 1st- to 3rd-graders (Thurstone's Primary Mental Abilities Test). Both tasks were presented on paper, and the experimenter recorded children's responses.

Children's Mental Transformation Task. Pre-K and kindergarten children were asked to determine which of four shapes would be formed by fitting two pieces (presented above the shapes) together (Levine, Huttenlocher, Taylor, & Langrock, 1999). Some items required mental rotation, and others required only mental translation. Participants completed 24 items, and mean accuracy was recorded (there were no practice trials on this task). We calculated internal reliability at T1 using McDonald's omega (ω = .73) (Dunn, Baguley, & Brunsden, 2014).

Thurstone's Primary Mental Abilities test, Mental Rotation subtest. Children in 1st to 4th grades were asked to choose one of four shapes, each rotated at different angles that would form a square when put together with the target shape (Thurstone, 1938). Before the task, the experimenter explained the definitions of square and rectangle, and children completed 4 practice items with feedback. Children completed 16 test trials without feedback, and mean accuracy was recorded (T1 reliability, ω = .57).

Table 1
Test–retest reliability across tasks.

Measure	Sample size per time point				Test–retest reliability		
	Time 1	Time 2	Time 3	Time 4	Time 1 to Time 2	Time 2 to Time 3	Time 3 to Time 4
VSWM	435	457	453	429	.45	.38	.44
Mental transformation	501	487	443	429	.57	.53	.56
Mental rotation	507	500	464	441	.52	.61	.61
Proportional reasoning	499	489	461	422	.43	.41	.49
AMS acuity	387	405	393	379	.47	.51	.55
Number line estimation	444	455	433	388	.58	.61	.66
Approximate symbolic calculation	449	461	432	391	.32	.31	.39
Exact calculation	465	451	405	399	.61	.53	.61
Reading achievement	461	460	423	407	.84	.82	.90
Verbal IQ	454	NA	NA	NA	NA	NA	NA

Note. Sample size and test–retest reliability for each task (total N = 612) are shown. All test–retest reliabilities were significantly different from zero (p < .001). Number line estimation and proportional reasoning percentage absolute error scores were arcsine transformed. VSWM, visuospatial working memory; AMS, analog magnitude system; NA, not applicable (assessed only at Time 1).

Mental rotation

We gave different age-appropriate mental rotation tasks to pre-K and kindergarten children (Ghost Puzzle) than to 1st- to 4th-graders (Letter Rotation). Both tasks were presented on paper, and responses were recorded by the experimenter.

Ghost Puzzle task. In this task, pre-K and kindergarten children were asked to mentally rotate two ghosts to determine which would fit into a circle with a hole in the middle that was the same shape as one of the ghosts (Frick, Hansen, & Newcombe, 2013). On each trial, the choices were the correct answer and its mirror image, rotated at different angles with respect to the hole. Before the test items, children completed 2 practice trials with feedback (the first involved manipulating foam board pieces to fit into a hole, and the second involved mentally rotating the same images). Participants then completed 22 test trials without feedback, and mean accuracy was calculated (T1 reliability, ω = .76).

Letter Rotation task. In this task, 1st- to 4th-graders were shown a target letter in a box on the left and four rotated letters on the right (Quaiser-Pohl, Neuburger, Heil, Jansen, & Schmelter, 2014). Children were instructed to decide which two of the four rotated letters were identical to the one on the left when it was spun right side up; the two incorrect letters were mirror-reversed. Prior to the test trials, participants were given two worked examples and 2 practice items with feedback. Participants then completed 16 test trials without feedback. A trial was scored as correct only if children chose both correct letters and no others. Mean accuracy was recorded (T1 reliability, ω = .88).

Proportional reasoning

We assessed proportional reasoning at all grade levels using a computerized task that has been previously used in 4- to 10-year-olds (Möhring et al., 2015; Möhring, Newcombe, Levine, & Frick, 2016). Participants were told a story about a bear named Paul who likes to drink mixtures of cherry juice and water. Children saw a vertical rectangle ("cup") divided into blue ("water") and red ("cherry juice") sections. The total height of the rectangle, as well as the proportion of red and blue segments, varied between trials. Participants were asked to click on a horizontal line below the cup to indicate how much each drink would taste like cherry on a scale from one cherry (not at all like cherries) on the left to many cherries (a lot like cherries) on the right. Children completed 2 practice trials with feedback and then completed 15 test items without feedback; item order was randomized across participants. We excluded participants' scores if 90% of their responses fell within 10% of the line's length, following prior work on removing repetitive responses from number line estimation data (Slusser et al., 2013). We calculated the percentage absolute error (PAE) between the target and the response for each item (PAE = | response − correct | / total line range) and then calculated an average PAE score for each participant (T1 reliability, ω = .87). We then reverse-scored PAE so that higher scores would indicate better performance.

Analog magnitude system acuity

Precision of children's AMS was assessed using Panamath software (Halberda, Mazzocco, & Feigenson, 2008). In this computerized task, children saw blue dots on one side of the screen and yellow dots on the other and were asked which color had more dots. Children were instructed that more dots meant a larger number of dots regardless of the size of the dots. Participants completed 1 practice trial on paper with feedback. For all children, half the items were size-controlled and half were not, and the number of dots per side ranged from 5 to 21. The number of items, display time, and ratios differed according to children's grade level.[2] We examined percentage accuracy because this is a more stable and reliable measure than the Weber fraction (Inglis & Gilmore, 2014). Published reliability with children and adults is from .69 to .74 (Spearman–Brown corrected random split-half reliability; Libertus, Odic, Feigenson, & Halberda, 2016; Libertus, Odic, & Halberda, 2012). In addition, work with adults

[2] Children in pre-K and kindergarten completed 48 items with a display time of 2128 ms and ratios from 1.23 to 3.00. Children in 1st and 2nd grades completed 56 items with a display time of 1789 ms and ratios from 1.20 to 2.86. Children in 3rd and 4th grades in Year 1 completed 56 items with a display time of 1506 ms and ratios from 1.17 to 2.71, and those in Year 2 completed 64 items with a display time of 1269 ms and ratios from 1.17 to 2.67.

suggests that dot discrimination tasks like Panamath are the most reliable measures of AMS acuity (Chesney, Bjalkebring, & Peters, 2015).

Number line estimation

Number line estimation was assessed using a computerized, bounded number-to-position task with labeled end points (Fazio, Bailey, Thompson, & Siegler, 2014; Siegler & Opfer, 2003). Pre-K and kindergarten children completed a 0–100 number line task, and 1st- to 3rd-graders completed the 0–1000 number line task, appropriate for assessing individual differences at these ages (Gunderson et al., 2012). Each task had 18 items presented in a randomized order (see Appendix B for all items).[3] There were no practice trials. For each item, children clicked on the line to indicate where a specific number should be positioned. The experimenter asked children not to spend too long thinking about each number in an effort to encourage estimation. Consistent with prior research (Slusser et al., 2013), we excluded participants' scores on this task if 90% of their responses fell within 10% of the line's length. We chose to use mean PAE in our analyses because it is a widely used, theory-neutral method of examining accuracy on the task. Reliabilities at T1 were ω_{0-100} = .77 and ω_{0-1000} = .87. We reverse-scored PAE so that higher scores would indicate better performance on all tasks.

Approximate symbolic calculation

Children completed a computerized approximate symbolic calculation task adapted from Gilmore et al. (2007). See Appendix B for all items. On each trial, children saw two characters, Sarah and John, and the experimenter explained that each character would receive some marbles, which were represented by bags labeled with Arabic numerals (see Fig. 2). On addition trials, one character received two bags of marbles consecutively. On subtraction trials, one character received a bag of marbles and then lost a bag of marbles. On all trials, the other character then received one bag of marbles. Children's task was to say which character had more marbles. The correct answer differed from the comparison total by one of two ratios (pre-K and kindergarten: 4:7 and 2:3; 1st and 2nd grades: 2:3 and 4:5; 3rd and 4th grades: 4:5 and 6:7). Addition items and subtraction items were blocked and counterbalanced. Each block contained 2 practice trials with feedback, followed by 10 test trials (4 easier ratio, 4 harder ratio, and 2 check trials described below). This resulted in 16 approximate calculation test trials (8 addition and 8 subtraction).[4] We examined percentage accuracy on these 16 items in our analyses (T1 reliabilities were calculated within grade-level groups, $\omega_{\text{pre-K and kindergarten}}$ = .67, $\omega_{\text{1st and 2nd grades}}$ = .56, and $\omega_{\text{3rd grade}}$ = .55).

We employed several strategies to encourage children to approximate on this task; the bags of marbles were occluded after being briefly presented, experimenters encouraged children to choose quickly, and the numbers represented in each trial varied by grade level and were chosen to be more difficult than children could typically complete exactly. We also included check trials (2 per block) in which the characters' total number of marbles differed by exactly one. Children did not perform significantly above chance on these check trials, based on one-sample *t* tests within grade and time point with correction for multiple comparisons (mean accuracy ranged from 43.4% to 55.5%). Chance-level performance on check trials indicates that, as a group, children were not able to perform these calculations exactly.

Exact calculation

To assess exact calculation, children in pre-K and kindergarten completed the story problems task, and children in 1st to 4th grades completed the Woodcock–Johnson IV Calculation subtest.

[3] At T1, some children received a version of the task with different numbers or more trials. We included data from these children. Details can be found in Appendix A.

[4] Due to an error in stimulus design, 3rd- and 4th-graders completed 7 approximate addition trials (3 6:7 ratio items and 4 4:5 ratio items) and 3 exact addition check trials. The 3rd- and 4th-graders completed the correct number of subtraction trials of each type (8 approximate subtraction items and 2 exact subtraction check trials). For each item type, we included all available items of that type in our analyses.

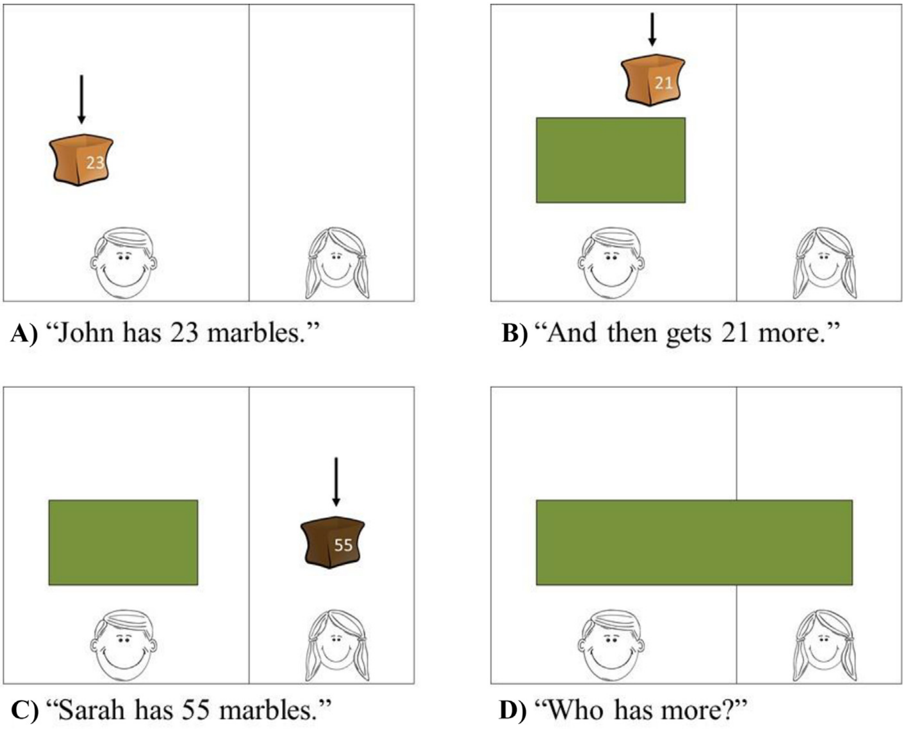

Fig. 2. Sample trial from approximate symbolic calculation task. Images were animated and appeared sequentially (in order from A to D). Arrows were not shown to participants but are included in the figure to indicate direction of motion in the animated sequence. (Adapted from C. K. Gilmore, S. E. McCarthy, & E. S. Spelke, 2007, "Symbolic arithmetic knowledge without instruction," *Nature, 447,* p. 589. Copyright 2007 by Nature Publishing Group.)

Story problems. Pre-K and kindergarten children completed an orally presented story problems task adapted from Levine, Jordan, and Huttenlocher (1992). The experimenter read each problem to children and recorded their verbal response. A total of 14 problems were presented in a single pseudorandom order (7 addition and 7 subtraction; see Appendix B). Operands and solutions ranged from 1 to 7; for example, "Eric has 4 balloons. Tiffany takes away 2 of his balloons. How many balloons does Eric have left?" Mean accuracy was recorded (T1 reliability, ω = .84).

Woodcock–Johnson IV Calculation. In this nationally normed paper-and-pencil test, 1st- to 4th-graders were asked to complete written arithmetic problems of increasing difficulty (Schrank, Mather, & McGrew, 2014). The items typically included addition, subtraction, multiplication, and division involving whole numbers and fractions. Testing continued until both basal criteria (6 lowest-numbered items administered were correct) and ceiling criteria (6 highest-numbered items administered were incorrect) were met. Scores were excluded if the experimenter discontinued testing before administering enough items to meet these basal and ceiling criteria (a type of experimenter error). Published reliability for 5- to 10-year-olds is greater than .90 (McGrew, LaForte, & Schrank, 2014). We analyzed children's W scores, which are Rasch-scaled scores suitable for assessing growth over time (Woodcock, 1999).

Reading achievement

As a control measure, we assessed reading achievement using the Woodcock–Johnson IV Letter–Word Identification subtest (Schrank et al., 2014). In this nationally normed test, children must identify individual letters and read words of increasing difficulty. We excluded children's scores if the basal and ceiling criteria were not met due to experimenter error. Reliability for 4- to 10-year-olds is greater than .94 (McGrew et al., 2014). We analyzed children's W scores on this test.

Verbal IQ

As a control measure, we assessed verbal IQ using the Kaufman Brief Intelligence Test–Second Edition (KBIT-2; Kaufman & Kaufman, 2004) at T1 only. Children were asked to point to one of six pictures that best matched a word or phrase and to provide a one-word answer to verbal riddles (e.g., "What is something that wags its tail and barks?" *dog*). The test contains items of increasing difficulty, and testing continued until basal and ceiling criteria were established. Published internal reliability on this measure for children aged 4–10 years is from .86 to .91 (Kaufman & Kaufman, 2004). We excluded children's scores on this task if basal and ceiling criteria were not met due to experimenter error. We examined standard scores based on age.

Analytic approach

All analyses were conducted in R (R Core Team, 2020). Prior to our analyses, we visually examined histograms of each measure. Because PAEs are expected to be non-normally distributed, we used an arcsine transformation on number line estimation and proportional reasoning PAE scores. All measures (including arcsine transformed PAEs) were approximately normally distributed with no evidence of floor or ceiling effects. In addition, we examined skewness and kurtosis as indicators of normality, with absolute values of skewness less than 2 and absolute values of kurtosis less than 7 considered acceptable (Curran, West, & Finch, 1996). All measures had values in the acceptable range. We did not exclude any values as outliers (after confirming that all values were valid scores on each measure) because values more than 3 standard deviations from the mean were expected given our large sample size.

Missing data

Missing data occurred for reasons such as attrition, child absence, child refusal, and experimenter error. For each task, a child's data were excluded if the child completed fewer than half the items on that task. There were no missing data for grade level, school, or time point, and there was 1 missing data point for child gender. For other variables, the percentage of missing data ranged from 16.7% to 36.8%. To examine the nature of the missing data, we performed Little's missing completely at random (MCAR) test, which was significantly different from zero, $\chi^2(5959) = 5720$, $p = .014$, indicating that the data were not MCAR. Therefore, for our main analyses, we used multiple imputation (MI) using the R package *mice* (van Buuren & Groothuis-Oudshoorn, 2011). MI reduces bias compared with listwise deletion, especially when data are not MCAR (van Ginkel, Linting, Rippe, & van der Voort, 2020). We used 20 imputations, the recommended number of imputations when the overall fraction of missing information is less than 30% (Graham, Olchowski, & Gilreath, 2007). Our MI model accounted for the nesting of longitudinal observations within participants. We used predictive mean matching, which allows for missing data to be imputed based on the observed data without making assumptions about the distribution of each variable (van Buuren, 2018).

Main analyses

Our main analyses examined the role of initial spatial skills in predicting children's initial levels and growth in exact calculation, approximate calculation, and number line estimation. We ran separate multilevel models for each numeracy outcome, with observations nested within children, using the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). In each model, the focal predictors were the five T1 spatial skills, which we entered as predictors of both the intercept (initial level) and the slope (growth rate) of the target numeracy skill. Because measures were standardized within grade level, a positive slope would indicate that a child improved relative to his or her grade-level peers on a given task. We entered time as a continuous variable based on each child's date of testing (where the intercept [time = 0] was the start date of the study) to account for variability in testing dates within a time point. In all models, we included a strict set of covariates (grade level, gender, school fixed effects, SES, initial reading achievement, and initial verbal IQ) as predictors of both the slope and intercept. We report pooled estimates from the multiply imputed datasets.

Results

Preliminary analyses

We conducted preliminary analyses on raw scores to establish the expected pattern of improvement with grade level. (Note that we only conducted these analyses among participants who completed the same task; e.g., comparing pre-K with kindergarten on the Ghost Puzzle Task, comparing 1st-graders with 2nd-graders on the Letter Rotation Task.) Grade level was significantly correlated with higher performance on all tasks (with the exception that T1 Panamath scores did not significantly differ between 1st and 2nd grades). See Appendix Table C1 for descriptive statistics of raw scores by grade level.

For several of our measures of spatial skills and numeracy, different tasks (or different levels of difficulty within a task) were administered to different grade levels. To analyze all participants' data on a single scale, we standardized children's scores by creating z scores within each grade level on each measure. These scores represent how many standard deviations each student scored above or below the mean score in their grade. These grade-standardized scores were positively correlated for all T1 spatial and numeric measures ($.15 < r < .44$; $ps < .01$; small to moderate effects[5]) (see Appendix Table C2 for correlations among key study variables). We confirmed that mental transformation and mental rotation were not highly related, $r(491) = .38$, $p < .001$, a moderate effect; therefore, we included these as separate predictors in our main analyses. We examined gender differences using t tests with Holm's adjustment for multiple comparisons (see Appendix Table C3). Boys performed significantly better than girls on number line estimation at T2, T3, and T4 ($.34 < d < .59$, $ps < .01$, small to moderate effects); there were no other significant gender differences.

Main analyses

Exact calculation

We first examined which specific initial spatial skills were most related to children's initial levels and growth rates in exact calculation. Model 1 (Table 2) included our five focal measures of initial spatial skills (VSWM, mental transformation, mental rotation, proportional reasoning, and AMS acuity) as predictors of initial levels and growth rates in exact calculation. We found that mental rotation, proportional reasoning, and AMS acuity each significantly contributed to children's initial levels of exact calculation skill, with small but significant effect sizes, in which a 1-*SD* increase in spatial skill was related to a .10- to .15-*SD* increase in initial exact calculation skill. However, contrary to our expectations, none of these spatial skills predicted growth rates in children's exact calculation skills. Furthermore, VSWM and mental transformation skills did not significantly relate to initial levels or growth rates in exact calculation skills.

We also asked whether number line estimation skill was a significant predictor of exact calculation over and above all the measured spatial skills and covariates (Table 2, Model 2). We found that initial number line estimation skill was a significant predictor of initial levels of exact calculation skill ($B = .17$, $SE = .05$, $p < .001$, a small effect size) but did not significantly predict growth in exact calculation skill.

Approximate calculation

We next examined the impact of initial spatial skills on approximate symbolic calculation levels and growth rates (Table 2, Model 3). In contrast to exact calculation skill, for approximate calculation

[5] We interpret correlation coefficients and standardized regression coefficients as follows. We interpret values of .10 to .30 as small, values greater than .30 to .50 as moderate, and values greater than .50 as large (Cohen, 1988; Peterson & Brown, 2005). We interpret Cohen's d values of .20 to .50 as small, values greater than .50 to .80 as moderate, and values greater than .80 as large (Cohen, 1988).

Table 2
Multilevel models predicting exact calculation, approximate calculation, and number line estimation from specific Time 1 spatial skills.

	Model 1 Exact Calculation *B* (*SE*)	Model 2 Exact Calculation *B* (*SE*)	Model 3 Approximate Calculation *B* (*SE*)	Model 4 Approximate Calculation *B* (*SE*)	Model 5 NL estimation *B* (*SE*)
Intercept					
VSWM	.05 (.05)	.03 (.05)	.07 (.05)	.06 (.05)	.09 (.05)
Mental transformation	.08 (.05)	.07 (.05)	.04 (.05)	.03 (.05)	.02 (.05)
Mental rotation	.11* (.05)	.10* (.05)	.07 (.05)	.06 (.05)	.08 (.05)
Proportional reasoning	.13** (.05)	.10* (.05)	.07 (.05)	.06 (.04)	.15*** (.04)
AMS acuity	.16** (.05)	.15** (.05)	.13** (.05)	.12* (.04)	.08 (.05)
NL estimation	–	.17*** (.05)	–	.11* (.05)	–
Slope					
VSWM	.01 (.04)	.01 (.04)	.03 (.05)	.02 (.05)	−.02 (.04)
Mental transformation	−.04 (.04)	−.04 (.04)	.00 (.05)	−.01 (.05)	.03 (.05)
Mental rotation	−.01 (.05)	.00 (.05)	−.02 (.05)	−.03 (.05)	−.02 (.05)
Proportional reasoning	−.03 (.04)	−.02 (.04)	−.01 (.04)	−.02 (.05)	−.04 (.04)
AMS acuity	−.03 (.05)	−.02 (.05)	−.01 (.04)	−.01 (.05)	.00 (.04)
NL estimation	–	−.06 (.04)	–	.03 (.05)	–
Variance components					
Intercept	.184	.162	.143	.127	.247
Slope	.089	.087	.069	.065	.099
Residual	.595	.595	.735	.738	.536
Variance explained in intercept (pseudo-R^2)	61.1%	65.8%	46.2%	52.3%	47.7%
Variance explained in slope (pseudo-R^2)	34.1%	35.6%	−11.3%	−4.8%	12.4%

Note. School fixed effects, starting grade level, gender, socioeconomic status, reading achievement, and verbal IQ were included as predictors of the slope and intercept (for coefficients of covariates, see Appendix Table C4). Time is treated as a continuous variable in years from September 1, 2015 (the start of the study). Each model included random effects on intercept and time. Pseudo-R^2 was calculated based on the change in each variance component compared with the null model (note that negative values are possible with this approach; Kreft & De Leeuw, 1998). NL, number line; VSWM, visuospatial working memory; AMS, analog magnitude system.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

skill only one spatial measure was a significant predictor of the intercept: AMS acuity ($B = .13$, $SE = .05$, $p = .009$, a small effect size). Again, no spatial skills predicted growth in approximate calculation.

In Model 4, we tested whether number line estimation was related to approximate calculation over and above other measured spatial skills. Here, we found that number line estimation was a significant predictor of the initial level (intercept) of approximate symbolic calculation ($B = .11$, $SE = .05$, $p = .029$, a small effect size) but did not significantly predict growth rates.

Comparing relations of number line estimation with exact versus approximate calculation. We had initially predicted that T1 number line estimation would be more strongly related to approximate calculation than to exact calculation. Unexpectedly, number line estimation significantly predicted initial approximate calculation (Model 4: $B = .11$, $SE = .05$, $p = .029$) and initial exact calculation (Model 2: $B = .17$, $SE = .05$, $p < .001$), with a directionally larger effect on exact calculation than on approximate calculation. To formally test whether these effects differed, we conducted an additional model with calculation type (exact vs. approximate) as an additional random effect. In this model, we found a non-significant interaction of number line estimation and calculation type predicting children's initial levels of calculation skill ($B = -.06$, $SE = .06$, $p = .368$). This indicates that the relation of number line estimation to calculation skill did not significantly differ between exact and approximate calculation tasks.

Number line estimation

We next examined whether each spatial skill would directly predict levels and growth rates in number line estimation (Table 2, Model 5). Only one spatial skill was significantly related to children's initial number line estimation: nonsymbolic proportional reasoning (B = .15, SE = .04, p < .001, a small effect size). No spatial skills significantly predicted growth in number line estimation.

Mediation

Our original longitudinal mediation hypotheses rested on our prediction that specific spatial skills would predict growth rates in number line estimation and exact calculation. Because this was not the case, we did not further pursue longitudinal mediation analyses.

Exploratory analyses

Moderation by grade

To test whether the impact of spatial skills on numeracy was moderated by grade level, we ran Models 1–5 with the addition of interaction terms for grade with each spatial skill predicting both the intercept and slope of the dependent variable. Using the Bonferroni correction for multiple comparisons (p < .01 considered significant), no grade-level interactions were significant (ps > .038).

Moderation by gender

We also re-ran Models 1–5, adding interaction terms for gender with each spatial skill as predictors of the intercept and slope of the dependent variable. No interaction terms were significant (ps > .05).

Single spatial factor score

To test whether collinearity among our five spatial measures (despite their low to moderate zero-order correlations) might have reduced our ability to detect relations of initial spatial skills to growth in numeracy skill, we also conducted Models 1–5 replacing the five spatial skills with a single spatial factor score (Appendix D). As in our main analyses, in all models the single spatial factor significantly predicted initial levels of symbolic numeracy (.22 < B < .33, ps < .001, small to moderate effect sizes) but not growth rates (ps > .05).

Discussion

In this 2-year cohort-sequential study of children in pre-K to 4th grade (4–10 years of age), we investigated three key questions: (1) whether specific spatial skills (VSWM, mental transformation, mental rotation, proportional reasoning, and AMS acuity) were differentially related to specific symbolic numeracy outcomes (approximate calculation and exact calculation), (2) whether spatial skills would predict initial levels or growth in symbolic numeracy skills, and (3) whether number line estimation skill would mediate these relations. By including five spatial skills, typically studied separately, we were able to parse apart the individual contributions of each spatial skill to symbolic numeracy. In doing so, we found unique relations between initial levels of specific spatial skills and specific numeracy skills, consistent with the *time-invariant hypothesis.* However, we did not find evidence that initial spatial skills predicted growth in any numeracy measure and therefore did not further investigate the possibility of the number line mediating such relations. We also found that number line estimation related to initial levels of both exact and approximate calculation skills but not growth in those skills. Thus, our findings did not support the *spatial–numeric growth hypothesis.* We discuss each key finding in turn.

Specific relations between spatial skills and numeracy skills

The first major contribution of this study was to reveal evidence of specificity regarding which spatial skills are most related to each aspect of numeracy. However, the specific pattern of relations only partially aligned with our predictions. We predicted that VSWM and either mental rotation or mental transformation would be most related to exact calculation, and we expected weak or no relations between proportional reasoning and AMS acuity and exact calculation. Consistent with this, mental rotation (but not mental transformation) was a significant predictor of initial levels of exact calculation skill. This aligns with the theory that mental rotation skills support the spatial mental models needed to learn exact arithmetic calculation skills. It is worth noting that we included both mental transformation and mental rotation tasks in our analyses, and only mental rotation significantly related to exact calculation skill. A key difference between these tasks is that, whereas mental transformation tasks can be solved using either mental rotation or alternative strategies such as feature matching, mental rotation tasks eliminate featural differences and therefore can be solved only using mental rotation. Thus, the fact that mental rotation, and not mental transformation, was related to exact calculation further suggests that the use of alternative strategies such as feature matching might not be related to calculation and that mental rotation per se is involved in exact calculation skill.

However, inconsistent with our predictions, VSWM was not a significant predictor of exact calculation skills after controlling for other spatial skills. This suggests that in prior studies linking VSWM to exact calculation in young children, VSWM may have served as a proxy for more directly relevant spatial skills such as mental rotation. This possibility is consistent with findings from previous work demonstrating that spatial and mathematics skills form separate unidimensional constructs and that VSWM maps onto the spatial factor (Mix et al., 2016, 2017). In addition, we found two unexpected relations between spatial skills and exact calculation: both nonsymbolic proportional reasoning skill and AMS acuity significantly related to children's initial levels of exact calculation skill with effect sizes similar to mental rotation. Although we had proposed that AMS acuity would be less related to exact tasks than to approximate tasks, this relation is nevertheless consistent with prior work showing that AMS acuity relates to math achievement at this age (Chen & Li, 2014). This finding adds to prior research by showing that the relation of AMS acuity to math achievement holds even after accounting for multiple spatial skills. We also found a significant unexpected relation between proportional reasoning and exact calculation. It is possible that the abstract relational thinking required for proportional reasoning is helpful for acquiring exact calculation skills or vice versa. Future research is needed to probe this relation further.

Regarding spatial predictors of approximate symbolic calculation, we had hypothesized that each of our spatial predictors would relate to approximate symbolic calculation (as mediated by number line estimation, discussed further below). Our results confirmed that approximate symbolic calculation was related to both AMS acuity and number line estimation. However, we did not find significant relations between approximate symbolic calculation and other measured spatial skills (VSWM, mental transformation, mental rotation, or proportional reasoning). This result is consistent with prior research and theory suggesting that approximate symbolic calculation is supported by a mapping of number symbols to analog magnitudes (e.g., Gilmore et al., 2007, 2014; Waring & Penner-Wilger, 2017). Our findings further show that both precision of the underlying magnitude representation (measured by the AMS task) and precision of the nonsymbolic-to-symbolic mapping (measured by the number line estimation task) are robust and specific correlates of this approximate symbolic calculation, after controlling for a variety of spatial skills.

We also hypothesized that each spatial skill would uniquely relate to number line estimation. However, we found that after accounting for all other spatial skills and covariates, only nonsymbolic proportional reasoning significantly related to number line estimation. This suggests that prior findings of relations between mental transformation and number line estimation may have been driven by children's proportional reasoning skills, which were not measured in these prior studies (Gunderson et al., 2012; LeFevre et al., 2013). Nonsymbolic proportional reasoning requires estimation of line lengths and relational reasoning. Thus, nonsymbolic proportional reasoning may be a key precursor to number line estimation, which involves these same skills and further requires symbolic number magnitude knowledge.

Together, these results indicate that specific spatial skills show unique relations to specific numeracy skills. Indeed, each numeracy skill was related to a distinct subset of spatial skills, and no single spatial skill predicted all numeracy skills. Proportional reasoning, a skill for which individual differences have not been well studied, emerged as a significant predictor of both exact calculation and number line estimation. These findings build on a recent factor analysis of spatial and mathematics skills, which found that a proportional matching task loaded on a mathematics factor for 6th-graders (Mix et al., 2017). Notably, Mix et al. (2017) used a forced-choice proportion matching task, which differs in response format from the task in the current study where proportions needed to be estimated on a continuous response scale. Children are generally more successful when reasoning about proportions continuously, whereas matching tasks (especially when they involve discretized proportions) can be biased by numerical or matching strategies (Begolli, Booth, Holmes, & Newcombe, 2020; Boyer, Levine, & Huttenlocher, 2008; Schlottmann, 2001; Spinillo & Bryant, 1999). Thus, our results may better capture early variability in proportional reasoning to provide further evidence that proportional reasoning contributes to children's mathematics development.

Why might proportional reasoning represent an important link between spatial skills and numeracy? Nonsymbolic proportional reasoning may provide an effective strategy that helps children to spatialize their numerical magnitude representations. Proportional reasoning involves reasoning about relations between spatial magnitudes such as line lengths, which may serve as a crucial precursor to number line estimation, which explicitly links numerical magnitudes and line lengths. Other spatial tasks like spatial scaling similarly require the ability to encode relative distances within different-sized spaces, a skill highly related to proportional reasoning (Möhring et al., 2015; Newcombe, Möhring, & Frick, 2018). Thus, proportional reasoning may provide children with tools for reasoning about mathematical concepts as well as other spatial concepts such as spatial scaling (Gilligan et al., 2019). These results suggest that proportional reasoning warrants further attention in research on spatial skills and numeracy and may be a useful target for intervention.

Spatial skills and number line estimation did not predict growth in symbolic numeracy

Contrary to our expectations, we did not find evidence that initial levels of spatial skills predicted growth in symbolic numeracy. Although these findings differ from a number of studies showing longitudinal relations between earlier spatial skills and later symbolic numeracy skills (e.g., Frick, 2019; Gunderson et al., 2012; LeFevre et al., 2013; Verdine et al., 2017), they are more consistent with studies that use similar analytic approaches to separate time-varying and time-invariant relations between spatial skills and numeracy (Bailey, 2017; Geer et al., 2019).

These results suggest that the relation between spatial and numerical skills is already stable across time at this age. This is sometimes described as "trait-level" stability (e.g., Bailey, 2017; Bailey, Watts, Littlefield, & Geary, 2014). Several theories are consistent with this type of stable relation. One possibility is that the relation between spatial skills and numeracy can be explained by a stable environmental confound. In other words, if children are in similar environments (e.g., home and school) across the pre-K to 4th-grade age range (4–10 years), and the environments or experiences that enhance spatial skills also tend to enhance numerical skills, this would lead to the observed pattern of data. Given that we controlled for children's verbal IQ, this stable environmental factor is unlikely to be explained by general cognitive stimulation and instead implicates a specific relation between environmental support for numerical skills and support for spatial skills. Indeed, parents who engage in more numerical activities at home also engage in more spatial activities (Zippert & Rittle-Johnson, 2020), suggesting that correlated environmental factors supporting both spatial and numerical skills could explain these results.

A second possibility is that the stable relations between spatial skills and numeracy are explained by a biological confound that remains stable across this age range. Specifically, the relation between spatial and numerical skills could be accounted for by individual differences in the strength or efficiency of shared neural circuits that are known to underlie both spatial and numerical skills (Hawes, Sokolowski, Ononye, & Ansari, 2019). It is important to note that individual differences in these shared neural mechanisms could develop as a result of either early environmental factors (prior to pre-K), inborn genetic factors, or their interactions.

A third potential explanation is that spatial and numerical skills are in fact causally related but that the impact of spatial skills on numeracy development has already occurred prior to pre-K (4–5 years of age), the earliest age in this study. If children's very early spatial skills caused advances in early numerical skills, this would lead to correlated individual differences in spatial and numerical skills that may remain stable across the observed age range.

Relation of number line estimation to exact and approximate calculation

We also found that children's initial number line estimation skills were significantly related to initial levels of exact and approximate calculation even after accounting for children's spatial skills. Although we had predicted that number line estimation would be more strongly related to approximate calculation than to exact calculation, our findings did not support this prediction. Rather, children's number line estimation skill was significantly related to both exact calculation skill (B = .17, p < .001) and approximate calculation skill (B = .11, p = .029), and these relations did not significantly differ from each other. This is consistent with prior research showing that number line estimation is related to many math achievement measures, including calculation (Schneider et al., 2018). It is also consistent with theories that children must recruit arithmetic competencies to solve traditional number line estimation tasks (Link, Nuerk, & Moeller, 2014).

We had also predicted that initial number line estimation skill would predict growth in approximate calculation skill. However, number line estimation was robustly related to initial numeracy skills but not to growth in numeracy skills. This is similar to our findings for spatial skills, and the same three explanations (environmental confounds, biological confounds, and an early causal relation) could explain this time-invariant relation between number line estimation and calculation skills.

Limitations and future directions

We note several limitations of the current study. First, we selected a broad set of spatial skills that we expected to relate to number line estimation. However, several spatial skills that have been associated with math skill were unmeasured in this study, including spatial scaling (Gilligan et al., 2019), visuomotor integration (Kurdek & Sinclair, 2001), disembedding (Mazzocco & Myers, 2003), and perspective taking (Guay & McDaniel, 1977). It is possible that one or more of these skills would also have related to children's numeracy. Nevertheless, the current study represents a step toward a comprehensive model of spatial–numeric relations across multiple spatial and numerical skills.

Second, for several spatial skills of interest, there was no single measure available that was appropriate for use across our target age range. Therefore, we administered age-appropriate measures that varied across grade levels. This leads to some limitations. Because scores were not directly comparable across grade levels, we were only able to model students' growth rates relative to their grade-level peers rather than in absolute terms. Future research would benefit from the development of normed spatial skill measures that can be used across ages 4–10 years.

Finally, a major limitation of the current study is its correlational design, which cannot definitively prove (or disprove) causal relations between measured variables. We attempted to mitigate this by using a robust set of covariates (grade level, gender, school fixed effects, SES, initial reading achievement, and initial verbal IQ) and by using a robust statistical approach that allowed us to disentangle the impact of initial spatial skills on initial levels and growth rates of numerical skills. Our analytic approach also allowed us to examine each spatial skill as a unique independent predictor after controlling for all other spatial skills and covariates. Nevertheless, to fully understand the relation between spatial skills and numeracy, it is important for researchers to use designs that allow for causal inference. Experimental interventions designed to strengthen spatial skills and investigate whether spatial skill improvements transfer to symbolic numeracy skills represent an important approach. Our research suggests that such interventions may be fruitful if they target preschool-aged children or earlier. Indeed, recent intervention studies found that spatial skills training was beneficial for both spatial and numerical skills in preschool as well as during middle childhood (Hawes et al., 2017; Lowrie, Harris, Logan, & Hegarty, 2021).

Conclusion

This 2-year longitudinal study using a large diverse sample of children in pre-K to 4th grade (4–10 years of age) provides novel evidence regarding the specificity and time course of relations between children's spatial and numerical skills. We investigated five distinct spatial skills that have been identified in prior literature as predictors of symbolic numeracy. Two of those skills, VSWM and mental transformation, did not account for unique variance in any of our symbolic numeracy skills after accounting for other spatial skills. The other three spatial skills showed unique patterns of relation to exact calculation (related to mental rotation, proportional reasoning, and AMS acuity), approximate calculation (related to AMS acuity only), and number line estimation (related to proportional reasoning only). Together, these results indicate specificity in the relations between specific numerical and spatial skills and suggest that researchers and educators should treat spatial skills as a multidimensional construct.

Notably, children's initial spatial skills and number line estimation skills were related only to their initial levels of symbolic numeracy and did not predict growth in symbolic numeracy, indicating that longitudinal correlations that do not account for time-invariant individual differences should be interpreted with caution. Nevertheless, time-invariant relations between spatial and numerical skills could reflect early-developing causal relations or environmental impacts on both types of skills. These possibilities warrant further research, especially given the importance of setting children onto positive trajectories in both numerical and spatial domains.

Acknowledgments

This research was supported by a National Science Foundation CAREER DRL (Division of Research on Learning) award (1452000) and a James S. McDonnell Foundation Scholar Award (220020546) to Elizabeth Gunderson. We thank the children and teachers who gave their time to this research and the full-time research assistants who managed data collection and data cleaning: Victoria Bartek, Conner Black, Jorge Carvalho Pereira, and Ying Lin. We also thank the undergraduate research assistants who helped carry it out: Synclaire Arther, Jennifer Bradley, Riley Brown, Yesha Dave, Olivia Dermody, Brittany Dumproff, John Durison, Lillian Ham, Maya Johnson, Jiai (Jenny) Jung, Christin Kim, Nicole Lee, Shuo (Lexi) Liu, Samantha McLaughlin, Victoria McLaughlin, Sumaiya Nusrath, Jessica Palmarini, Gabriella Riccardo, Sherly Smith, Jae Sung, Audrey Wrobel, and Najah Young.

Appendices A-D. Supplementary Materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2021.105251.

References

Alloway, T. P. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment.
Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.). *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.
Bailey, D. H. (2017). Causal inference and the spatial–math link in early childhood. *Monographs of the Society for Research in Child Development, 82*(1), 127–136. https://doi.org/10.1111/mono.12288.
Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science, 25*(11), 2017–2026. https://doi.org/10.1177/0956797614547539.
Bakker, M., Torbeyns, J., Wijns, N., Verschaffel, L., & De Smedt, B. (2019). Gender equality in 4- to 5-year-old preschoolers' early numerical competencies. *Developmental Science, 22*(1), e12718. https://doi.org/10.1111/desc.12718.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Begolli, K. N., Booth, J. L., Holmes, C. A., & Newcombe, N. S. (2020). How many apples make a quarter? The challenge of discrete proportional formats. *Journal of Experimental Child Psychology, 192*, 104774. https://doi.org/10.1016/j.jecp.2019.104774.
Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology, 114*(3), 375–388. https://doi.org/10.1016/j.jecp.2012.09.015.
Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*(4), 1016–1031. https://doi.org/10.1111/j.1467-8624.2008.01173.x.
Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. *Developmental Psychology, 44*(5), 1478–1490. https://doi.org/10.1037/a0013110.
Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition, 128*(3), 331–352. https://doi.org/10.1016/j.cognition.2013.04.008.
Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test-scores across diverse samples. *Developmental Psychology, 31*(4), 697–705. https://doi.org/10.1037/0012-1649.31.4.697.

Caviola, S., Mammarella, I. C., Cornoldi, C., & Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *Journal of Experimental Child Psychology, 112*(2), 141–160. https://doi.org/10.1016/j.jecp.2012.02.005.
Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica, 148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016.
Chen, Z., & Siegler, R. S. (2000). II. Overlapping waves theory. *Monographs of the Society for Research in Child Development, 65*(2), 7–11. https://doi.org/10.1111/1540-5834.00075.
Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186.
Chesney, D., Bjalkebring, P., & Peters, E. (2015). How to estimate how well people estimate: Evaluating measures of individual differences in the approximate number system. *Attention, Perception, & Psychophysics, 77*(8), 2781–2802. https://doi.org/10.3758/s13414-015-0974-6.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. https://doi.org/10.1037/1082-989X.1.1.16.
De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001.
Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7*(4), 145–147. https://doi.org/10.1016/S1364-6613(03)00055-X.
Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. https://doi.org/10.1111/bjop.12046.
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.
Foley, A. E., Vasilyeva, M., & Laski, E. V. (2017). Children's use of decomposition strategies mediates the visuospatial memory and arithmetic accuracy relation. *British Journal of Developmental Psychology, 35*(2), 303–309. https://doi.org/10.1111/bjdp.12166.
Frick, A. (2019). Spatial transformation abilities and their relation to later mathematics performance. *Psychological Research Psychologische Forschung, 83*(7), 1465–1484. https://doi.org/10.1007/s00426-018-1008-5.
Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development, 28*(4), 386–399. https://doi.org/10.1016/j.cogdev.2013.06.002.
Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology, 55*(3), 637–652. https://doi.org/10.1037/dev0000649. 10.1037/dev0000649.supp (Supplemental).
Georges, C., Cornu, V., & Schiltz, C. (2019). Spatial skills first: The importance of mental rotation for arithmetic skill acquisition. *Journal of Numerical Cognition, 5*(1), 5–23. https://doi.org/10.5964/jnc.v5i1.165.
Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2019). The developmental relations between spatial cognition and mathematics in primary school children. *Developmental Science, 22*(4). https://doi.org/10.1111/desc.12786.
Gilmore, C. K., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: Exploring the relationships among different tasks. *Learning and Individual Differences, 29*, 50–58 https://doi.org/10.1016/j.lindif.2013.10.004.
Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447*, 589–592. https://doi.org/10.1038/nature05850.
Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention science, 8*(3), 206–213. https://doi.org/10.1007/s11121-007-0070-9.
Guay, R. B., & McDaniel, E. D. (1977). The relationship between mathematics achievement and spatial abilities among elementary school children. *Journal for Research in Mathematics Education, 8*(3), 211–215. https://doi.org/10.2307/748522.
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.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. https://doi.org/10.1037/a0012682.
Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665–669. https://doi.org/10.1038/nature07246.
Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cognition and Instruction, 35*(3), 236–264. https://doi.org/10.1080/07370008.2017.1323902.
Hawes, Z., Moss, J., Caswell, B., Seo, J., & Ansari, D. (2019). Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach. *Cognitive Psychology, 109*, 68–90. https://doi.org/10.1016/j.cogpsych.2018.12.002.
Hawes, Z., Sokolowski, H. M., Ononye, C. B., & Ansari, D. (2019). Neural underpinnings of numerical and spatial cognition: An fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation. *Neuroscience & Biobehavioral Reviews, 103*, 316–336. https://doi.org/10.1016/j.neubiorev.2019.05.007.
Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684.
Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review, 107*(3), 500–524. https://doi.org/10.1037/0033-295X.107.3.500.

Hutchison, J. E., Lyons, I. M., & Ansari, D. (2019). More similar than different: Gender differences in children's basic numerical skills are the exception not the rule. *Child Development, 90*(1), e66–e79. https://doi.org/10.1111/cdev.13044.
Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review, 18*(6), 1222–1229. https://doi.org/10.3758/s13423-011-0154-1.
Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica, 145*(Supplement C), 147–155. https://doi.org/10.1016/j.actpsy.2013.11.009.
Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test* (2nd ed.). Pearson Inc.
Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Science of Learning, 3*(1), 12. https://doi.org/10.1038/s41539-018-0028-7.
Kreft, I. G., & De Leeuw, J. (1998). Introducing Multilevel Modeling. *SAGE Publications, Ltd.*. https://doi.org/10.4135/9781849209366.
Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*(3), 451–455. https://doi.org/10.1037/0022-0663.93.3.451.
Kyttälä, M., Aunio, P., Lehto, J. E., Van Luit, J. E. H., & Hautamäki, J. (2003). Visuospatial working memory and early numeracy. *Educational and Child Psychology, 20*(3), 65–76.
Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*(3), 195–216. https://doi.org/10.1016/j.lindif.2005.12.001.
Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: A meta-analytic review. *Psychological Bulletin*. https://doi.org/10.1037/bul0000191. 10.1037/bul0000191.supp (Supplemental).
LeFevre, J.-A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S.-L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00641.
Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35*(4), 940–949. https://doi.org/10.1037/0012-1649.35.4.940.
Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology, 53*(1), 72–103. https://doi.org/10.1016/S0022-0965(05)80005-0.
Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*. https://doi.org/10.1111/j.1467-7687.2011.01080.x.
Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *Journal of Experimental Child Psychology, 150*, 207–226. https://doi.org/10.1016/j.jecp.2016.06.003.
Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*(3), 373–379. https://doi.org/10.1016/j.actpsy.2012.09.009.
Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology, 67*(8), 1597–1613. https://doi.org/10.1080/17470218.2014.892517.
Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2014). The association between children's numerical magnitude processing and mental multi-digit subtraction. *Acta Psychologica, 145*, 75–83. https://doi.org/10.1016/j.actpsy.2013.10.008.
Lowrie, T., Harris, D., Logan, T., & Hegarty, M. (2021). The impact of a spatial intervention program on students' spatial reasoning and mathematics performance. *The Journal of Experimental Education, 89*(2), 259–277. https://doi.org/10.1080/00220973.2019.1684869.
Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia, 53*(1), 218–253. https://doi.org/10.1007/s11881-003-0011-7.
McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Woodcock-Johnson IV Technical Manual. Riverside.
McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visuospatial interference on children's arithmetical performance. *Educational and Child Psychology, 20*(3), 93–108.
Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences, 20*(2), 101–109. https://doi.org/10.1016/j.lindif.2009.08.004.
Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences, 26*, 141–152. https://doi.org/10.1016/j.lindif.2012.03.012.
Mix, K. S. (2019). Why are spatial skill and mathematics related? *Child Development Perspectives, 13*(2), 121–126. https://doi.org/10.1111/cdep.12323.
Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: developmental and educational implications. In J. B. Benson (Ed.). *Advances in child development and behavior* (Vol. 42, pp. 197–243). Elsevier.
Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D., & Bower, C. (2021). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology, 113*(2), 304–314. https://doi.org/10.1037/edu0000494. 10.1037/edu0000494.supp (Supplemental).
Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145*(9), 1206–1227. https://doi.org/10.1037/xge0000182. 10.1037/xge0000182.supp (Supplemental).
Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C. J., Hambrick, D. Z., & Konstantopoulos, S. (2017). The latent structure of spatial skills and mathematics: A replication of the two-factor model. *Journal of Cognition and Development, 1–28*. https://doi.org/10.1080/15248372.2017.1346658.
Möhring, W., Newcombe, N. S., & Frick, A. (2014). Zooming in on spatial scaling: Preschool children and adults use mental transformations to scale spaces. *Developmental Psychology, 50*(5), 1614–1619. https://doi.org/10.1037/a0035905.
Möhring, W., Newcombe, N. S., & Frick, A. (2015). The relation between spatial thinking and proportional reasoning in preschoolers. *Journal of Experimental Child Psychology, 132*, 213–220. https://doi.org/10.1016/j.jecp.2015.01.005.

Möhring, W., Newcombe, N. S., Levine, S. C., & Frick, A. (2016). Spatial proportional reasoning is associated with formal knowledge about fractions. *Journal of Cognition and Development, 17*(1), 67–84. https://doi.org/10.1080/15248372.2014.996289.

National Governors Association Center for Best Practices Council of Chief State School Officers. (2010). Common Core State Standards. National Governors Association Center for Best Practices, Council of Chief State School Officers.

Newcombe, N. S., Möhring, W., & Frick, A. (2018). Chapter 9 - How big is many? Development of spatial and numerical magnitude understanding. In A. Henik & W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition* (pp. 157–176). Academic Press.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science, 297*(5587), 1708–1711. https://doi.org/10.1126/science.1072493.

Nys, J., & Content, A. (2010). Complex mental arithmetic: The contribution of the number sense. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 64*(3), 215–220. https://doi.org/10.1037/a0020767.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction, 23*, 125–135. https://doi.org/10.1016/j.learninstruc.2012.08.004.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175–181. https://doi.org/10.1037/0021-9010.90.1.175.

Piaget, J., & Inhelder, B. (1975). *The origins of the idea of chance in children*. Norton.

Quaiser-Pohl, C., Neuburger, S., Heil, M., Jansen, P., & Schmelter, A. (2014). Is the male advantage in mental-rotation performance task independent? On the usability of chronometric tests and paper-and-pencil tests in children. *International Journal of Testing, 14*(2), 122–142. https://doi.org/10.1080/15305058.2013.860148.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005.

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.

Ren, K., Lin, Y., & Gunderson, E. A. (2019). The role of inhibitory control in strategy change: The case of linear measurement. *Development Psycology, 55*(7), 1389–1399. https://doi.org/10.1037/dev0000739.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*(4), 525–536. https://doi.org/10.1111/desc.12166.

Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72*(1), 103–122 http://www.jstor.org/stable/1132474.

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

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.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV Tests of Achievement. Riverside.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–702.

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691–697. https://doi.org/10.1177/0956797612440101.

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives, 8*(3), 144–150. https://doi.org/10.1111/cdep.12077.

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

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

Singer-Freeman, K. E., & Goswami, U. (2001). Does half a pizza equal half a box of chocolates?: Proportional matching in an analogy task. *Cognitive Development, 16*(3), 811–829. https://doi.org/10.1016/S0885-2014(01)00066-1.

Skagerlund, K., & Träff, U. (2016). Processing of space, time, and number contributes to mathematical abilities above and beyond domain-general cognitive abilities. *Journal of Experimental Child Psychology, 143*, 85–101. https://doi.org/10.1016/j.jecp.2015.10.016.

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.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? *A critical review. American Psychologist, 60*(9), 950–958. https://doi.org/10.1037/0003-066x.60.9.950.

Spelke, E. S., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition, 78*, 45–88. https://doi.org/10.1016/S0010-0277(00)00108-6.

Spinillo, A. G., & Bryant, P. E. (1999). Proportional reasoning in young children: Part-part comparisons about continuous and discontinuous quantity [Article]. *Mathematical Cognition, 5*(2), 181–197. https://doi.org/10.1080/135467999387298.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1302751110.

Thurstone, L. L. (1938). *Primary Mental Abilities*. University of Chicago Press.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? *Psychology of Learning and Motivation, 57*, 147–181. https://doi.org/10.1016/B978-0-12-394293-7.00004-2.
Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402. https://doi.org/10.1037/a0028446.
van Buuren, S. (2018). *Flexible Imputation of Missing Data* (2nd ed.). CRC/Chapman & Hall.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03.
van Ginkel, J. R., Linting, M., Rippe, R. C. A., & van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of Personality Assessment, 102*(3), 297–308. https://doi.org/10.1080/00223891.2018.1530680.
Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Spatial skills, their development, and their links to mathematics. *Monographs of the Society for Research in Child Development, 82*(1), 7–30. https://doi.org/10.1111/mono.12280.
Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835. https://doi.org/10.1037/a0016127.
Waring, R. J., & Penner-Wilger, M. (2017). Estimation of importance: Relative contributions of symbolic and non-symbolic number systems to exact and approximate calculation. *Journal of Numerical Cognition, 2*(3), 202–219. https://doi.org/10.5964/jnc.v2i3.9.
Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360. https://doi.org/10.3102/0013189x14553660.
Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–127). Erlbaum.
Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., & van Lieshout, E. C. D. M. (2013). Individual differences in kindergarten math achievement: The integrative roles of approximation skills and working memory. *Learning and Individual Differences, 28*, 119–129. https://doi.org/10.1016/j.lindif.2013.09.012.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1–B11. https://doi.org/10.1016/S0010-0277(99)00066-9.
Zippert, E. L., & Rittle-Johnson, B. (2020). The home math environment: More than numeracy. *Early Childhood Research Quarterly, 50*, 4–15. https://doi.org/10.1016/j.ecresq.2018.07.009.