

Children Underperform Following “Math” but Not “Spatial” Task Framing

Lindsey Hildebrand and Sara Cordes

Department of Psychology and Neuroscience, Boston College

Increasing evidence suggests that success in science, technology, engineering, and math (STEM) fields is not only dependent upon one’s actual STEM-relevant abilities but also upon one’s STEM-relevant attitudes—in particular, math and spatial attitudes. Here, we examine whether simply mentioning the math or spatial relevance of a task affects children’s performance and the moderating role of children’s math and spatial attitudes. Further, we examine gender differences in performance given pervasive gender gaps in STEM and early-emerging gender differences in math and spatial attitudes. Participants (221 first- to fourth-grade children from the United States; 113 girls, 108 boys; 52% White, 16% Black, 14% Asian, 9% Hispanic or Latinx, 18% multiple races/ethnicities) were introduced to a novel task framed as tapping into math or spatial abilities (or no framing [control condition]). Children then completed math and spatial anxiety and self-concept measures. Results indicate that children who heard the math task framing were less accurate relative to children in the control condition, and the effect was larger for those with higher math anxiety or lower math self-concept, but it was not different for boys and girls. Children who heard the spatial task framing, however, performed comparably to children in the control condition. Though both math and spatial attitudes revealed identical patterns of gender differences (with higher anxiety and lower self-concept in girls than boys), there were no gender differences in performance. This study highlights the salient role of math attitudes early in development and provides key insights for future work aimed at increasing STEM outcomes.

Public Significance Statement

Findings reveal that children perform worse on a novel task when it is described as being related to math—but not spatial—skills, and this effect is strongest for children with more negative math attitudes. Moreover, girls had more negative math and spatial attitudes than did boys, but gender differences were not observed in performance on the task. Findings provide key insights into processes implicated in science, technology, engineering, and math outcomes.

Keywords: math attitudes; spatial attitudes; math anxiety; math self-concept; science, technology, engineering, and math outcomes

Supplemental materials: <https://doi.org/10.1037/dev0001603.suppl>

Growing the science, technology, engineering, and mathematics (STEM) workforce has robust benefits, both for individuals and for society as a whole. A STEM career can provide individuals with a pathway to economic security: The average annual salary for a STEM occupation is nearly 35% greater than that of a non-STEM occupation (\$95,000 vs. \$72,000; [National Center for Science and Engineering Statistics, 2021](#)) for individuals with at least a college degree. More broadly, to address pressing issues of global

importance such as pandemics and climate change, it is necessary to have a large, highly skilled STEM workforce.

Despite the importance of STEM, numerous factors have been identified as barriers to the pursuit of and success in STEM careers, many of which have roots in early childhood. Two important factors highlighted as critically important for later STEM outcomes are attitudes toward STEM and STEM-related domains (e.g., [Wang et al., 2013, 2015](#)) and gender ([Ceci et al., 2009](#)). In particular, negative

This study was not preregistered. This work was supported by National

have no known conflicts of interest to disclose. Data Lindsey Hildebrand

This article was published Online First November 2, 2023, and materials are available on the Open Science Framework: <https://osf.io/> Lindsey Hildebrand <https://orcid.org/0000-0003-3632-7083> y9bmh/.

Science Foundation Grant DRL-1920732 to Sara Cordes. We thank the members of the Boston College Infant and Child Cognition Lab, especially Lauren Sprague, Lesenia Fish, Hayley Liebenow, Celine Lim, Sophia Lee, and Caroline Kraus, who assisted in data collection and management. Thank you to Katherine McAuliffe and David Miele for their feedback on the manuscript. We also extend our gratitude to the Acton Discovery Museum, the Boston Children’s Museum, the Living Laboratory at the Museum of Science in Boston, and all of the families who participated in this study. We

served as lead for conceptualization, data curation, formal analysis, investigation, methodology, visualization, and writing—original draft. Sara Cordes served as lead for funding acquisition and supervision and served in a supporting role for conceptualization, investigation, and methodology. Lindsey Hildebrand and Sara Cordes contributed equally to project administration and writing—review and editing.

Correspondence concerning this article should be addressed to Lindsey Hildebrand, Department of Psychology and Neuroscience, Boston College, 140 Commonwealth Avenue, 300 McGuinn Hall, Chestnut Hill, MA 02467, United States. Email: hildebrandlindsey@gmail.com

math and spatial attitudes are related to lower performance on math and spatial tasks (e.g., Ramirez et al., 2012, 2013), and gender differences in these attitudes have been documented, such that women and girls report higher math and spatial anxieties and lower self-concepts than men and boys (e.g., Alvarez-Vargas et al., 2020; Cvencek et al., 2011; Lauer et al., 2018). Additionally, as early as first grade, children endorse math and spatial gender stereotypes, associating these domains with boys and men (e.g., Cvencek et al., 2011; Moè, 2018). However, though emerging work suggests math and spatial attitudes are related in adulthood (Daker et al., 2022; Delage et al., 2021; Sokolowski et al., 2019), how math and spatial attitudes compare in childhood, and how gender differences in these attitudes vary across childhood, remain open questions. Further, little is known how and when these attitudes predict outcomes in STEM contexts early in development.

In this study, we explored the concurrent relation between math and spatial attitudes and the moderating role of these attitudes for performance on a novel task. We focused on understanding a potential circumstance that impacts performance—task-relevant framing. Is it something about the task demands, or alternatively, the child's beliefs about the task demands, that drives domain-relevant attitudes or beliefs to impact performance on a task? To explore this question, we presented first to fourth-grade children with a task framed as pertaining to math or spatial skills. We assessed their performance (i.e., accuracy) relative to a baseline condition (with no framing) and explored how performance on the task varied as a function of math and spatial attitudes. Further, given the salient role of gender in STEM contexts, we paid particular attention to age-related variation in gender differences in math and spatial attitudes and gender differences in performance on the framed task. In this cross-sectional study of children across early childhood (i.e., first to fourth grades), we aimed to: (a) assess whether children's performance on a task is related to framing the task as "math" or "spatial" in nature, (b) determine whether children's math or spatial attitudes and/or gender moderate any observed performance differences, and (c) characterize gender differences in math and spatial attitudes in first to fourth graders.

Math and Spatial Attitudes Math Attitudes

Math anxiety is one of the most well-studied math attitudes,¹ broadly defined as tension, worry, or fear associated with thinking about or performing math tasks (Ashcraft, 2002; Hembree, 1990). Along with math self-concept—defined as one's perception of their math ability (Bong & Clark, 1999)—both have been implicated as critical for long-term STEM outcomes (Ahmed et al., 2012). Both high math anxiety and low math self-concept are linked to lower math performance in childhood (e.g., Arens et al., 2022; Kriegbaum et al., 2015; Vukovic et al., 2013) and adulthood (e.g.,

Ashcraft & Kirk, 2001). Meta-analyses including work with adults and children have found a small-to-moderate relation between math anxiety and math performance (Barroso et al., 2021), and a moderate relation between math self-concept and math

performance (Lee, 2009).

It is well-documented that women report higher math anxiety and lower math self-concept than men (e.g., Dowker et al., 2016; Hart & Ganley, 2019; for meta-analysis, see Hyde et al., 1990), but there is mixed evidence of how early in development these gender differences in math attitudes emerge. Gender differences in math self-concept as early as 6 years of age have been consistently documented in early childhood through adulthood (e.g., Ahmed et al., 2012; Cvencek et al., 2011; Lee, 2009). However, early-emerging gender differences in math anxiety are less consistently found. For example, a few studies have found no difference in boys' and girls' math anxiety in childhood (Harari et al., 2013; Jameson, 2014; Ramirez et al., 2013), though other studies have reported gender differences in math anxiety around 6 years of age (Ahmed et al., 2012; Gunderson et al., 2017; Hill et al., 2016; Lauer et al., 2018; Lee, 2009). Thus, the evidence for gender differences in math anxiety (but not self-concept) in childhood appears to be less robust.

Spatial Attitudes

Recent work has suggested that spatial attitudes may also be relevant for STEM success (Daker et al., 2022; Delage et al., 2021; Sokolowski et al., 2019). As expected, the handful of studies investigating spatial anxiety have revealed negative relations between spatial anxiety and performance on spatial tasks such as mental rotation, both in adults and elementary-school-aged children (Alvarez-Vargas et al., 2020; Lauer et al., 2018; Ramirez et al., 2012). Work with adults has found consistent relations between higher spatial self-concept and higher spatial performance (Blajenkova et al., 2006). However, no work to date has explored spatial self-concept in children.

Similar to the domain of math, there are well-documented gender differences in spatial attitudes. Women consistently report higher spatial anxiety and lower spatial self-concept than men (Alvarez-Vargas et al., 2020; Hegarty et al., 2002; Lyons et al., 2018; Sokolowski et al., 2019). There is some evidence that gender differences in spatial anxiety may exist in elementary-school-aged children as well (Lauer et al., 2018; Ramirez et al., 2012; but not in preschool, see Wong, 2017). Critically, only a few studies have examined children's spatial anxiety with most of these studies investigating narrow age ranges (i.e., only including children from a 1- to 2-year age range; e.g., Ramirez et al., 2012; Wong, 2017) making it difficult to determine when and how gender differences in spatial anxiety may vary over the course of early childhood. Moreover, whether gender differences in spatial self-concept exist in childhood is an open question. Critically, emerging work supports the potential that math and spatial anxieties are unique but contribute to cross-domain outcomes (Daker et al., 2022; Delage et al., 2021; Sokolowski et al., 2019). For example, recent work has found an explanatory role of spatial anxiety for gender differences

¹ Though some researchers classify math anxiety as related to affect or emotion rather than as an attitude (e.g., Hembree, 1990), this distinction is primarily a theoretical divide in defining the construct, rather than one of an

empirical nature. For the purposes of the current article, we refer to math (and spatial) anxiety as an attitude.

in math anxiety in adulthood (Delage et al., 2021; Sokolowski et al., 2019), highlighting the importance of studying math and spatial attitudes concurrently.

In the current study, we characterize math and spatial attitudes in a cross-sectional sample of children across early childhood (first to fourth grades; approximately 6–10-years-old in the United States). We examine gender differences in these attitudes and how gender differences may vary as a function of grade. Moreover, we explore whether children who hear a novel task labeled as pertaining to “math” or “spatial” skills perform differently from those in a control condition (with no task label), and critically whether domainspecific attitudes may moderate observed performance differences. We also consider the role of working memory in the relation between math and spatial attitudes and performance. Working memory has been identified in both the math and spatial domains as a factor contributing to the relation between anxiety and performance (e.g., Lauer et al., 2018; Ramirez et al., 2012, 2013), potentially because anxiety co-opts working memory resources necessary for problemsolving (Ashcraft & Krause, 2007). As such, past work has found the relation between anxiety and performance is strongest for children (Lauer et al., 2018; Ramirez et al., 2012, 2013) and adults (Ashcraft & Krause, 2007) with higher working memory. Therefore, in the current study, we also consider the potential role of working memory as a moderator of framing effects on performance.

The Role of Gender in Math and Spatial Domains

Previous work has found that adults associate math and spatial domains more with males than females (e.g., Hirnstein et al., 2014; Nosek et al., 2002). Though findings from work with children are less consistent, math-male and spatial-male stereotypes have been found in children as young as age 6 (Cvencek et al., 2011; Hildebrand et al., 2022; Hirnstein et al., 2014; Neuburger et al., 2015; Nosek et al., 2002; though see Martinot et al., 2012; Moè, 2018). Whether math and spatial gender stereotypes are dominant in childhood or not, it is imperative to understand whether boys and girls respond differently in STEM-relevant contexts, possibly as a consequence of these stereotypes. For example, stereotypes may inhibit girls’ success in STEM contexts because they communicate those tasks as difficult or not appropriate for girls (Master, 2021). In the current work, we examine whether we find gender differences in performance following math or spatial task framing on a novel task. If observed, this could suggest that STEM-related gender stereotypes are influential in early childhood and can be induced even in novel contexts.

Task Framing and Performance

In this study, we explored how simply describing a task as pertaining to math or spatial skills may impact task performance in children. On the surface, highlighting the category associated with a task (e.g., calling something a “math task”) may not seem problematic. In fact, one may argue that task framing could even be beneficial such that an individual could bring to mind relevant knowledge that would help them on the task or could consider the broader relevance of the task. However, there is both empirical and theoretical support that task framing may have a noticeable negative

impact on performance outcomes (e.g., Cimpian, 2010; Cimpian et al., 2007, 2012). In one example, research has found that when generic descriptions about a domain are used (e.g., “smart people are good at math”), children—particularly those with high math anxiety—exhibit poorer performance, motivation, and attitudes, as compared to children who hear nongeneric descriptions of the same domain (“some smart people are good at math”; for reviews, see Dweck, 2007a, 2007b). In the current work, we explore how framing a task as relevant to math or spatial domains impacts performance on that task, and, importantly, we focus on two distinct factors that could be linked to task framing and performance: gender and domainspecific anxieties.

The Current Study

In the current study, we examine whether framing a novel task as related to math or spatial skills impacts performance on the task. We then investigate whether the following key factors moderate framing effects: math and spatial attitudes, working memory, and/or gender. In the process, we also provide one of the first characterizations of math and spatial anxieties and self-concepts in early childhood within the same study, with a specific focus on gender differences.

To examine relations between math and spatial task framing and performance, we presented first to fourth-grade students—our youngest participants are children who are just beginning to receive formal math instruction—with a novel task framed as a “math game,” a “spatial game,” or just as a “new game” (control). The task, a nonsymbolic magnitude comparison task, required children to indicate which of two simultaneously presented arrays of dots had more items. We specifically chose this task as it was both unfamiliar to children, and also had meaningful connections to math and spatial skills. Indeed, both numerical and spatial elements of the task impact accuracy (DeWind et al., 2015; Halberda et al., 2008), and in fact, performance on this task has been linked to other measures of math and spatial skill (e.g., Carr et al., 2020; Gunderson & Hildebrand, 2021; Halberda et al., 2008; Libertus et al., 2011).

We hypothesized that framing the task as pertaining to math or spatial abilities would influence performance relative to a baseline condition (no framing), with worse performance in some children, and (possibly) better performance in others. If simply highlighting the math or spatial nature of a task is related to decrements in a child’s performance on the task, this finding would have profound implications for understanding influences on STEM outcomes and direct implications for pedagogy.

Further, it was hypothesized that mathframing effects would lead to the largest differences in performance (relative to baseline), while spatial framing effects would be weaker or potentially nonexistent. This prediction stemmed from the fact that children experience math and spatial tasks in very distinct contexts. Children have formal experiences with graded assessments for math in the classroom, whereas their experiences with spatial tasks tend to be informal and involve playful activities (e.g., puzzles, building blocks; e.g., Dearing & Tang, 2010; Huntsinger et al., 2016; Jirout & Newcombe, 2015). As such, children’s more formal experiences with math may serve to heighten domain-specific anxieties and/or lead to more pronounced reactions to stereotypes than in comparison to spatial tasks.

Finally, we sought to test potential theoretical factors that moderate framing effects. We had two nonmutually exclusive hypotheses regarding potential moderators of the relation between task framing and performance:

Hypothesis 1: Given findings of gender differences in math and spatial attitudes (e.g., Hildebrand et al., 2022; Hyde et al., 1990; Lauer et al., 2018) and stereotypes that associate males with math and spatial tasks in childhood (e.g., Cvencek et al., 2011; Vander Heyden et al., 2016), it is possible that gender differences in performance will be found when the math or spatial nature of a task is made explicit. That is, girls may be more likely to underperform in the math and spatial framed conditions (relative to baseline), a potential consequence of current cultural stereotypes. In contrast, boys may perform better in the framed conditions (relative to baseline) given that math and space are domains in which their gender is stereotypically expected to excel. These results would be expected if task framing taps into children's gender stereotypes favoring boys in math and spatial domains.

Hypothesis 2: Alternatively, it is also possible that children's attitudes toward math and space may be particularly salient when performing a task framed as tapping into math or spatial abilities. Hearing that the task is related to math or spatial abilities may tap into individual attitudes toward these domains, which may heighten awareness of related anxiety and/or other attitudes and beliefs about one's own performance. This possibility is supported by work showing negative relations between math and spatial attitudes and performance outcomes for both males and females across the lifespan (e.g., Barroso et al., 2021; Lauer et al., 2018; Lyons et al., 2018; Ramirez et al., 2012). If so, children with higher math anxiety (and/or lower math self-concept) are predicted to perform worse on the task when it is framed as related to spatial abilities. Gender differences in performance are only expected inasmuch as they track with individual differences in domain-specific anxiety.

A second aim was to characterize gender differences in math and spatial anxiety and self-concept concurrently in early childhood. In line with prior work, we expected to find gender differences in math and spatial anxieties and self-concepts, with girls indicating more negative attitudes than boys. We did not have specific predictions regarding differences between children of different grades. For math attitudes, some past work has suggested that math attitudes are more negative in older grades given more formal experiences with math (e.g., Krinzinger et al., 2009). Alternatively, others have argued that math attitudes are similar across early childhood (e.g., Wang et al., 2020). Notably, this study extends previous work to compare math and spatial attitudes within the same children, allowing for a direct comparison of how attitudes in these two domains may compare in early childhood, a comparison with both theoretical and practical implications.

Method

Participants

Participants were 221 children in first to fourth grade ($M_{\text{age}} = 8.16$, $SD_{\text{age}} = 1.10$; 113 girls, 108 boys; for a detailed grade and gender breakdown, see [Supplement A in the online supplemental materials](#)) from an urban area in the Eastern United States. Data were collected in lab ($n = 52$), at local parks ($n = 34$), at local museums ($n = 55$), or at local schools ($n = 80$). We collected individual-level race and ethnicity for participants tested in lab ($n_{\text{race}} = 44$; 59% White, 14% Black, 5% Asian, 5% Hispanic, and 18% multiple races and/or ethnicities). For schools where data was collected, we computed the expected racial and ethnic composition of the sample based on the school-level composition and the number of students tested at each school (46% White, 16% Black, 11% Hispanic, 19% Asian, and 18% multiple races and/or ethnicities).

We used stratified random assignment to ensure an approximately equal number of boys and girls across condition and two-grade bins (first/second grade and third/fourth grade). An a priori power analysis determined a sample size of 206 was needed to detect medium effects ($f = .25$) for the factors: condition (three levels), gender (two levels), grade (two levels), and their interactions with a .90 power level and α set at .05 (Faul et al., 2007). An additional 11 children began the study but were excluded for the following reasons: child did not complete multiple tasks ($n = 4$), experimenter error ($n = 3$), computer error ($n = 3$), and parental interference ($n = 1$). Study procedures were approved by the Boston College

Institutional Review Board, Protocol 10.064.11.

Procedure

Participants completed the following tasks in this order: (a) math, spatial, or control framed task, (b) math and spatial anxiety questionnaires (order counterbalanced), (c) working memory task, and (d) math and spatial self-concept questionnaires (order counterbalanced). A subset of participants ($n = 97$) also completed a measure of generalized anxiety at the end of the session. Participants who received this assessment completed the measure as the final task.

Framed Task

Conditions. There were three between-subject conditions: math framing, spatial framing, and control (no framing). In the two framing conditions, the experimenter introduced the participant to the task by telling them they would be doing a "[math/spatial]" game, then described what it meant to be good at [math/spatial] games, and gave examples of [math/spatial] games. Following this explanation and instructions on how to perform the game, participants were told "Remember, this is a [math/spatial] game! People who are good at [math/spatial games] do well at this game" before starting the task. In the control condition, participants were

introduced to the task by hearing that they would be doing a “new” game but were given no additional information. Complete task framing scripts can be found in [Supplement A in the online supplemental materials](#).

Participants then completed the task, a computerized nonsymbolic numerical comparison task, in which participants were asked to repeatedly judge which of two arrays had more dots (Panamath; Halberda & Feigenson, 2008). In each trial, participants saw a grey screen with yellow dots on the left side and blue dots on the right side of the screen. These dots were presented for 600 ms after which a visual mask was flashed for 200 ms and then only the grey screen remained. As per Halberda and Feigenson (2008), the sizes of the dots were heterogeneous, with the average dot size equated across the two arrays on a random half of the trials, and the cumulative area of the dots equated across the two arrays on the other random half of trials. Although the task requires a relative numerosity judgment, prior work has shown that these spatial controls strongly influence performance on the task (Tokita & Ishiguchi, 2013). Furthermore, in the instructions, we excluded traditional numerical language and used language about size and “more” given that these aspects of the task apply to both math and spatial domains (for exact wording, see [Supplement A in the online supplemental materials](#)). Children indicated their response using keys that corresponded to the side of the screen where dots of that color appeared. The program settings for each participant were set to three minutes at medium difficulty. The program determined the size and ratio of the number of dots for each trial based on the participant’s age which was entered with the other settings. Based on their age, each participant completed 56 (6–7-year-olds) or 64 (8–10-year-olds) trials. Performance on this task was assessed based on accuracy.

Math and Spatial Anxiety

Children then completed the math and spatial anxiety scales (order counterbalanced). Following the framed task, children were told that they were now going to play some different games. Children were first introduced to the concept of nervousness (e.g., “some people are nervous when they don’t know an answer,” see Ramirez et al., 2013 for full script). Children were then introduced to a five-item response scale in which each option was represented by faces of varying levels of distress: a calm face on the far left (not nervous at all; scored as 1) and a very nervous face on the far right (very, very nervous; scored as 5). Children then completed a practice item about their anticipated anxiety on a task unrelated to math or spatial tasks to ensure comprehension in using the response scale (Ramirez et al., 2012, 2013). We used an average score in analyses.

To measure math anxiety, participants completed the eight-item Child Math Anxiety Questionnaire (CMAQ; Ramirez et al., 2013). Items on the CMAQ ask how anxious children would feel during specific math situations. The questionnaire contains two types of questions, situations that require (a) nonspecific math ability (e.g., “how anxious would you feel if you were called on by a teacher to explain a math problem on the board?”) and (b) specific math abilities (e.g., “how anxious would you feel if you had to solve the problem, ‘there are 13 ducks in the water, there are 6 ducks on land, how many ducks are there in all?’”). For items about specific math tasks, we adapted items to be more generic given that our age range

was bigger than the age range used during the construction of the CMAQ. For example, an item about specific math homework problems was changed to be about “hard math homework.” Reliability was acceptable ($\alpha = .76$). Of note, the reliability of our adaptation of the CMAQ was better than the reliability of the original measure ($\alpha = .55$; Ramirez et al., 2013).

To measure spatial anxiety, children completed the eight-item Children’s Spatial Anxiety Questionnaire (Ramirez et al., 2012). Children were asked to indicate how anxious they would feel during different activities that require spatial skills. Items included tasks like those requiring navigation (e.g., how to get from school to their house) and mental rotation (e.g., recreating a block structure from a picture). Reliability for the CMAQ was acceptable ($\alpha = .70$), which was again greater than during the test’s creation ($\alpha = .56$; Ramirez et al., 2012).

Working Memory

The working memory task was the Forward and Backwards Digit Span tests (from the Wechsler Intelligence Scale for Children; Wechsler, 2003). In light of previous findings that working memory may moderate the relation between math anxiety and math performance, this task was included to explore whether performance varied as a function of working memory in the framed conditions. Analyses did not reveal working memory to be a significant factor ($ps > .05$) and thus this measure is not discussed further. See the [online supplemental materials for detailed task description and analyses \(Supplement B\)](#).

Math and Spatial Self-Concept

Participants then completed the math and spatial self-concept measures (order counterbalanced; matched to order of anxiety questionnaires). To assess children’s self-concepts, we adapted the items from the math and spatial anxiety scales used. Rather than asking children to indicate their anxiety, we rephrased the items and asked children to indicate if they agreed that they were “good at” each task (item). We chose to use the same items from the anxiety scales so that we would have parallel measures of anxiety and self-concept, ensuring increased validity in comparisons between anxiety and self-concept. Children indicated their agreement with each statement via a 6-point Likert scale that ranged from disagree a lot to agree a lot. Use of this scale required children to first indicate whether they agreed or disagreed with the statement, and then whether they agreed/disagreed a little, kind of, or a lot (represented by circles of increasing size). This response format has been widely used with children (e.g., Cvencek et al., 2011; Gunderson et al., 2017). The two-part format has been found to reduce social desirability effects in this age range (Harter, 1982), and is shown to be more effective than parallel scales without visual representations of responses (Rebok et al., 2001). We used average scores in our analyses. Reliability was acceptable (math self-concept: $\alpha = .75$; spatial self-concept: $\alpha = .61$).

Generalized Anxiety Measure

A subset of participants ($n = 97$) completed the general anxiety subscale of the Screen for Child Anxiety Related Emotional

Disorders (Birmaher et al., 1997). This measure was added approximately halfway through data collection so that exploratory analyses could be run to examine whether effects related to math and spatial anxiety were unique to these domains, or instead, if they could be explained by patterns related to generalized anxiety (see [Supplement C in the online supplemental materials for full task description and analyses involving this measure in the online supplemental materials](#)).

Data Analyses

For all analyses, we report results using more conventional frequentist models as well as Bayesian models. Though Bayesian statistics offer many advantages as compared to frequentist statistics (for comprehensive overviews, see [Dienes & Mclatchie, 2018](#); [Kruschke & Liddell, 2018](#); [Wagenmakers et al., 2018](#)), their inclusion in the present work is particularly compelling in that it is possible to quantify the evidence both for and against a given effect. As such, Bayesian models provide a more comprehensive picture of reported effects and allow for direct comparisons of effects (e.g., gender differences in attitudes across domains and grades).

For Bayesian analyses, the notation B_{10} denotes the Bayes Factor (BF) for the alternate hypothesis (comparing the alternate model, 1, to the null model, 0) and B_{01} represents the BF for the null hypothesis. These values are the inverse of each other; however, we report the BF (either B_{10} or B_{01}) which is larger than one for ease of interpretation. Each BF can be interpreted as how much more likely the data are under the reported model (i.e., alternate or null) relative to the other model. For example, if $B_{10} = 8$, this indicates that the data are eight times more likely under the alternate model than the null model. Bayesian statistics do not use traditional significance cutoffs (i.e., $p < .05$), and instead simply consider the strength of the evidence for the alternate or null model. General interpretations of strength of the evidence based on the BF are as follows: $BF = 1$: no evidence; $BF = 1-3$: anecdotal evidence, $BF = 3-10$: substantial evidence; $BF = 10-30$: strong evidence; $BF = 30-100$: very strong evidence; $BF > 100$: decisive evidence ([Wagenmakers et al., 2011](#)). For models with multiple effects (e.g., age and gender), we report BF_{incl} or BF_{excl} for each effect which is the BF for the inclusion or exclusion of the effect in the model, respectively. Bayes analyses used default priors ([JASP Team, 2021](#)).

Planned Analyses

Framing Effects. To assess the impact of task framing on performance, we conducted an analysis of variance (ANOVA) examining group differences in performance based on the framing condition (control, space, or math), participant gender (boy or girl), and grade² (two-grade bins: first to second graders, third to fourth graders). Raw accuracy on the task was the dependent variable. All post hoc comparisons reflect Holm–Bonferroni corrections for multiple comparisons. This analysis also served to test whether framing effects varied as a function of child gender, addressing the

proposed hypothesis regarding the potential impact of gender stereotypes on performance.

Moderation Analyses. To test whether framing effects varied as a function of math and spatial attitudes, we used multiple regression models to test moderation effects. Specifically, moderation analyses examined whether math and spatial attitudes moderated the relation between task framing and performance. Separate models were run for (a) math anxiety, (b) math self-concept, (c) spatial anxiety, and (d) spatial self-concept. Given that we only expected attitudes to moderate condition effects within the same domain, we only included the framing condition for the same domain (e.g., math framing for models of math attitudes) and the control condition (see [Supplement D in the online supplemental materials for crossdomain moderation models in the online supplemental materials](#)). The control condition was always coded as the reference group. We used the overall analysis of framing effects to determine whether to include gender or age interactions in attitude moderation models. That is, if the effects of gender and grade were not significant in the overall model, we did not examine interaction effects related to gender and grade in the moderation models and instead included these factors as covariates.

For Bayesian moderation models, we report the statistics for the model including all factors. Per Bayesian model comparison, a Bayes Factor is calculated for each possible alternative model for the specified predictors in comparison to the null model. Thus, we additionally report the model that was determined as the best fit for the data. If none of the alternative models had evidence that was stronger than the evidence for the null model, we report the Bayes Factor for the null model.

Characterizing Math and Spatial Attitudes. To characterize math and spatial anxieties and self-concepts, we ran separate ANOVAs for each attitude measure examining the (a) main effect of gender (two levels), (b) the main effect of grade (two levels), and (c) the interaction between grade and gender. Additionally, condition was included in models given that the attitudinal measures were administered following the experimental framing (though results are the same if we do not include condition).

We report an exploratory analysis (independent samples *t* test) of gender differences in generalized anxiety to examine whether gender differences were present for anxiety broadly speaking, or if gender differences were unique to math and/or spatial anxiety. However, due to the small number of participants who completed this measure, we were underpowered to detect effects that were not very large ($f = .25$). Thus, we have included the description and discussion of this measure in the [online supplemental materials \(Supplement C\)](#).

Exclusion Criteria

In some cases, a participant's data were missing on a task due to experimenter error, computer error, or the child's choice not to complete the task. Twelve children (5.43%) had missing data on one or more task(s). If a child completed at least half of the items on a task, we included their data. Additionally, a participant's data from an individual task was excluded if it was greater than 1.5 times

² Because settings in our framed task varied as a function of the child's age, but our design accounted for grade, all analyses were rerun using age (in years), rather than grade. The pattern of findings using age was similar

with those reported here; thus, we report grade-related effects in line with the original aims of the study.

the interquartile range (IQR, i.e., the range of the middle 50% of the data) outside the 25th or 75th quartile (percentage of data removed across tasks: $M = 0.06\%$; range $0.00\%–3.65\%$; see Yang et al. (2019) for simulation demonstrating advantages of using IQR over standard deviation for outlier detection). We conducted a missing values analysis to test whether excluded data (missing and outlier) was Missing Completely at Random (MCAR). Little's MCAR test was not significant, $\chi^2(8, N = 221) = 8.48, p = .388$, suggesting data were not missing systematically. Thus, for each analysis, all participants were included who had data for each of the model variables (i.e., pairwise deletion).

Transparency and Openness

We report our a priori power analysis to determine sample size, all data exclusions, all manipulations, and all measures in the study, and we follow JARS (Kazak, 2018). Materials and data are available at <https://osf.io/y9bmh> (Hildebrand & Cordes, 2023). Data were analyzed using JASP (JASP Team, 2021). Figures were created in R, Version 4.0.0 (R Core Team, 2020), and used the packages ggplot, Version 3.2.1 (Wickham, 2016), and dplyr, Version 1.0.5 (Wickham et al., 2021). This study's design and analyses were not formally preregistered but were fully detailed in a funded National Science Foundation grant (DRL-1920732).

Results

Descriptive Statistics and Correlations

Descriptive statistics for all measures are reported in Table 1. For all correlations, we accounted for condition given that measures were administered following the experimental task framing. Partial correlations between all measures are reported in Table 2. Partial correlations between all measures split by child gender are presented in Table 3. In sum, we found that all math and spatial attitudes had moderate to high correlations. Additionally, even when partialing out variance due to condition, overall performance on the framed task was related to math anxiety, math self-concept, and spatial anxiety.

Table 1
Descriptive Statistics for All Measures Overall, by Gender, and by Condition

Measure	Possible range	All (N = 221) M (SD)	Girls (n = 113) M (SD)	Boys (n = 108) M (SD)	Test of gender differences d [95% CI]	Control condition (n = 74) M (SD)	Math condition (n = 73) M (SD)	Spatial condition (n = 74) M (SD)
Attitudes^a								
Math anxiety	1–5	2.56 (0.78)	2.74 (0.77)	2.37 (0.74)	0.48 [0.21, 0.75]	2.56 (0.79)	2.56 (0.78)	2.56 (0.78)
Spatial anxiety	1–5	2.42 (0.72)	2.61 (0.75)	2.22 (0.64)	0.53 [0.26, 0.79]	2.43 (0.76)	2.48 (0.73)	2.35 (0.67)
Math self-concept	1–6	4.31 (0.92)	4.02 (0.95)	4.61 (0.80)	0.66 [0.38, 0.93]	4.32 (0.90)	4.34 (0.95)	4.26 (0.98)
Spatial self-concept	1–6	4.38 (0.83)	4.10 (0.80)	4.67 (0.76)	0.69 [0.41, 0.96]	4.37 (0.89)	4.39 (0.81)	4.37 (0.79)
Performance								
Accuracy	0–100	91.87 (5.65)	91.45 (5.70)	92.30 (5.59)		93.14 (4.73)	89.99 (6.47)	92.69 (5.01)

Note. Effect sizes for attitude measures reflect that girls had more negative attitudes than boys (i.e., higher anxiety, lower self-concept). All tests of gender differences were significant, $p < .001$, BF_{10} range: 47.91–15ps, 090.51. CI. .30. =

Math and Spatial Framing Effects

Analyses revealed a main effect of condition on children's accuracy in the task and there was strong evidence in support of this effect, $F(2, 194) = 5.87, p = .003, \eta^2 = 0.06, BF_{incl} = 19.88$;

Figure 1. This main effect reflected significantly poorer performance of children in the math framing condition compared to the other two conditions and there was no difference in performance between the control and spatial framing conditions. There was strong evidence for the difference between the control and math conditions (mean difference = 2.95, 95% confidence interval, CI [0.73, 5.17], $p = .006, d = 0.52, BF_{10} = 22.30$), substantial evidence for the difference between the spatial and math conditions (mean difference = 2.57, [0.32, 4.82], $p = .015, d = 0.44, BF_{10} = 4.92$), and substantial evidence against a difference between the control and spatial conditions (mean difference = 0.38, [-1.93, 2.68], $p = .698, d = 0.08$,

$BF_{01} = 4.76$; Figure 1). Contrary to our expectations, the interaction between gender and condition was not significant and Bayesian analyses indicated substantial evidence against this interaction, $F(2, 194) = 0.63, p = .533, BF_{excl} = 7.52$, suggesting that performance was lower following the math framing but not the spatial framing for both girls and boys. The main effects of participant gender and grade were not significant ($ps > .30, BF_{excl} = 3.00$). Further,

Table 2 Correlations Between All Measures
confidence interval; BF = Bayes Factor.

Bayesian model comparisons indicated the best model of the data only included the effect of condition ($BF_{10} = 19.44$).

Moderation Analyses

As a reminder, moderation analyses examined how framing effects on performance varied as a function of each attitude,

only for the framing condition that matched the domain (e.g., math framing for models of math attitudes) and the control condition. Moderator variables were mean centered to aid in interpretation of parameter estimates.

Math Attitudes

Results of the math anxiety moderation model suggested that the main effect of condition was qualified by a significant

interaction with math anxiety (Figure 2a, Table 4). Post hoc analyses within condition revealed that performance was negatively related to math anxiety only in the math framing condition, $F(3, 69) = 5.85$, $p = .001$, $B = -3.67$, $\beta = -0.44$, $t = -4.04$, $p = .001$. In the control condition, there was no relation between anxiety and performance, $F(3, 68) = 1.54$, $p = .211$, $B = -0.56$, $\beta = -0.09$, $t = -0.77$, $p = 0.473$. Indeed, children with lower levels of math anxiety

Measure	1 r (BF ₁₀)	2 r (BF ₁₀)	3 r (BF ₁₀)	4 r (BF ₁₀)	5 r (BF ₁₀)
Demographics					
1. Grade (four levels)	—				
Attitudes					
2. Math anxiety	.24*** (13.01)	—			
3. Spatial anxiety	.34*** (21,069.96)	.61*** (3.18e + 19)	—		
4. Math self-concept	.20** (15.27)	.68*** (1.78e + 25)	.55*** (2.43e + 15)	—	
5. Spatial self-concept	.24*** (173.63)	.40*** (4.81e + 6)	.67*** (1.85e + 27)	.57*** (7.74e + 18)	—
Performance (framed task)					
6. Accuracy	.08 (0.19)	.29*** (212.61)	.16* (0.61)	.19** (3.68)	.10 (0.27)

Note. We report BF₁₀ for all correlations, regardless of whether it is greater than 1, for ease of presentation. Significant effects are bolded for ease of viewing. BF = Bayes Factor.

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

Table 3 Correlations Between All Measures, Split by Gender

Measure	1 r (BF ₁₀)	2 r (BF ₁₀)	3 r (BF ₁₀)	4 r (BF ₁₀)	5 r (BF ₁₀)	6 r (BF ₁₀)
Demographics						
1. Grade (four levels)	—	-.19 [†] (0.57)	.41*** (2,458.67)	.18 (0.98)	.24* (2.76)	.02 (0.12)
Attitudes						
2. Math anxiety.30*						(2.53)
3. Spatial anxiety.22*	.22** (3.63)	—	.56*** (7.10e + 6)	.64*** (2.28e + 10)	.37*** (118.65)	(0.40)
4. Math self-concept.25*	.24* (2.61)	.61*** (9.08e + 9)	—	.45*** (1.24e + 5)	.60*** (3.21e + 9)	(1.90)
5. Spatial self- (1.01)	.18 [†] (2.57)	.67*** (1.94e + 11)	.57*** (4.15e + 6)	—	.50*** (1.20e + 7)	concept.21*
Performance (framed task)						
6. Accuracy .13	.22* (11.14)	.31*** (43.57)	.70*** (8.26e + 13)	.53*** (9.15e + 6)		(0.41)
	.26** (8.10)	-.07 (0.18)	.10 (0.30)	-.03 (0.13)	—	

Note. Correlations for girls are above the dashes and correlations for boys are below the dashes and shaded. We report BF₁₀ for all correlations, regardless of whether it is greater than 1, for ease of presentation. Significant effects are bolded for ease of viewing. BF = Bayes Factor.

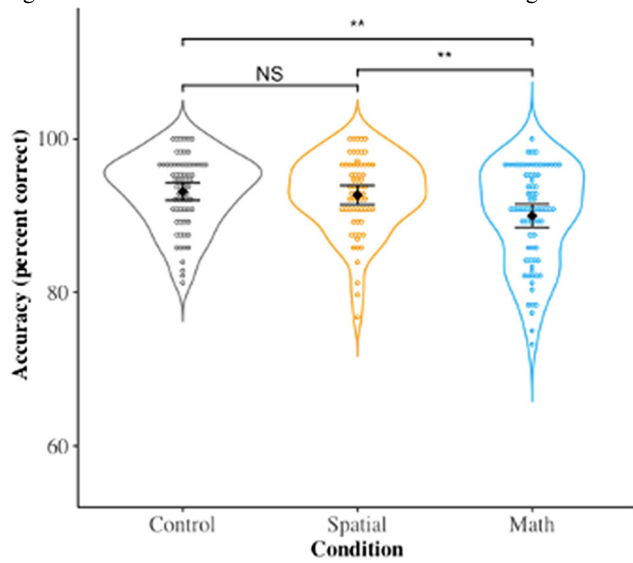
[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

(1 standard deviation below the mean) performed comparably across the two conditions, $t(35) = -0.38$, $p = .710$, $BF_{01} = 2.97$, whereas children with higher levels of math anxiety (1 standard deviation above the mean) performed worse in the math framing condition than the control condition, $t(26) = 3.16$, $p = .004$, $BF_{10} = 10.25$.

Similarly, math self-concept was a significant moderator of task framing (Figure 2b, Table 4). For children in the math condition, children with more positive math self-concept performed better, $F(3, 68) = 3.77$, $p = .015$, $B = 2.59$, $\beta = 0.36$, $t = 3.11$, $p = .003$,

whereas performance did not vary as a function of math self-concept in the control condition, $F(3, 62) = 0.88$, $p = .455$, $B = 0.07$, $\beta = 0.01$, $t = 0.09$, $p = 0.926$. For children with more positive math self-concept (1 standard deviation above the mean), performance was not significantly different between the control condition and math

Figure 1 Condition Differences Based on Task Framing



Note. Violin plots depict mean accuracy as well as individual data points by framing conditions. Error bars represent ± 1 standard error. Significance tests between conditions reflect Holm–Bonferroni comparisons. NS = not significant. See the online article for the color version of this figure.

** $p < .01$.

framing condition, $t(20) = 0.11$, $p = .911$, $BF_{01} = 2.59$. Paralleling results observed for math anxiety, children with the more negative math self-concept (1 standard deviation below the mean) had lower performance in the math condition relative to the control condition, $t(22) = 3.22$, $p = .004$, $BF_{10} = 10.36$. Notably, though the moderation factor was significant, the main effect of condition remained significant, indicating that math self-concept did not fully moderate the relation between framing and performance in the math framed condition.

Spatial Attitudes

Moderation models examining spatial anxiety and spatial self-concept indicated there was no relation between spatial anxiety or self-concept and performance, either overall or as a function of condition (Figure 2c and d, Table 4).

Characterizing Children's Math and Spatial Attitudes Math Anxiety

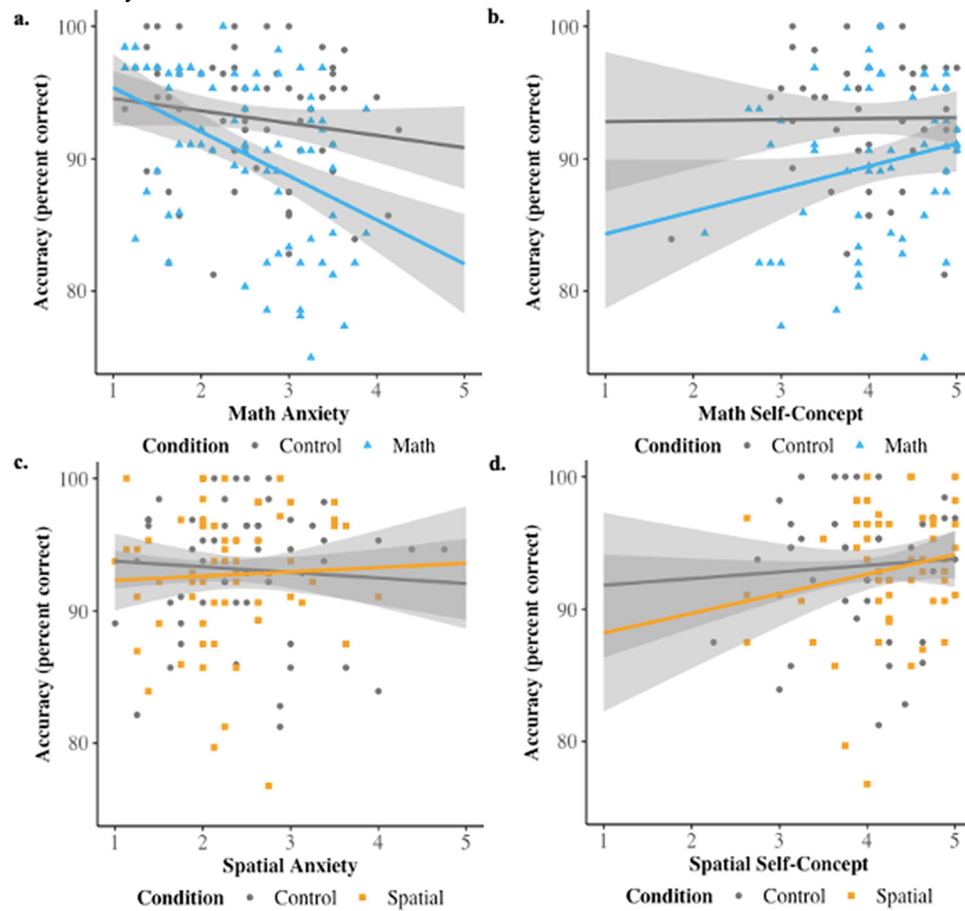
Children's math anxiety varied as a function of grade such that third and fourth graders indicated less math anxiety than children in first and second grades, $F(1, 213) = 7.31$, $p = .007$, $\eta^2 = 0.03$, $BF_{incl} = 5.13$; mean difference = 0.28, 95% CI [0.08, 0.49], $d = .38$. Consistent with previous work, girls reported higher math anxiety than boys, $F(1, 213) = 11.67$, $p = .001$, $\eta^2 = 0.05$, $BF_{incl} = 39.24$; mean difference = 0.35, 95% CI [0.15, 0.56], $d = .47$. The interaction between grade and gender was not significant, $F(1, 213) = 0.04$, $p =$

.837, $\eta^2 = 0.00$, $BF_{excl} = 5.68$, suggesting that gender differences were relatively similar across the two age groups. Condition did not account for significant variance, $F(2, 213) = 0.03$, $p = .967$, $\eta^2 = 0.00$, $BF_{excl} = 22.22$.

Math Self-Concept

Older children had more positive math self-concepts than younger children, $F(1, 209) = 7.16$, $p = .008$, $\eta^2 = 0.03$, $BF_{incl} = 4.51$, mean difference = -0.32 , 95% CI $[-0.55, -0.08]$, $d = .38$. Aligning with other findings, we found girls had more negative math self-concept than boys, $F(1, 209) = 21.03$, $p = .001$, $\eta^2 = 0.09$, $BF_{incl} = 3,536.13$, mean difference = -0.54 , 95% CI $[-0.77, -0.31]$, $d = .65$. These effects were not qualified by an interaction, $F(1, 209) = 0.16$, $p = .690$, $\eta^2 = 0.00$, $BF_{excl} = 4.74$. As expected,

Figure 2
Condition by Attitude Moderation on Children's Performance



Note. Trend lines reflect relations between performance and (a) math anxiety, (b) math self-concept, (c) spatial anxiety, and (d) spatial self-concept, by condition. The shaded area represents the 95% confidence interval around the trend line. See the online article for the color version of this figure.

the effect of condition was not significant, $F(2, 209) = 0.27$, $p = .761$, $\eta^2 = 0.00$, $BF_{\text{excl}} = 16.67$.

Spatial Anxiety

In line with effects observed with math anxiety, spatial anxiety varied as a function of grade and gender, grade:² $F(1, 213) = 22.02$, $p = .001$, $\eta^2 = 0.09$, $BF_{\text{incl}} = 2,784.65$, mean difference = 0.42, 95% CI [0.25, 0.60],² $d = 0.63$; gender: $F(1, 213) = 14.81$, $p = .001$, $\eta^2 = 0.06$, $BF_{\text{incl}} = 187.59$, mean difference = 0.35, [0.17, 0.53], $d = 0.53$, such that older children indicated lower levels of spatial anxiety and that girls had greater spatial anxiety than boys. As with math anxiety, the interaction between grade and gender was not significant, $F(1, 213) = 0.64$, $p = .426$, $\eta^2 = 0.00$,

$BF_{\text{excl}} = 3.98$, nor was the effect of condition, $F(2, 213) = 1.24$, $p = .291$, $\eta^2 = 0.01$, $BF_{\text{excl}} = 6.99$.

Spatial Self-Concept

Findings with spatial self-concept again mirrored findings observed for math such that the main effects of grade and gender were significant but the interaction between these effects was not, nor was the effect of condition, Grade: $F(1, 209) = 9.94$, $p = .002$, $\eta^2 = 0.04$, $BF_{\text{incl}} = 16.28$, mean difference = -0.33, 95% CI [-0.54, -0.12], $d = 0.44$; gender: $F(1, 209) = 24.73$, $p = .001$, $\eta^2 = 0.10$, $BF_{\text{incl}} = 12,953.75$, mean difference = -0.52, [-0.73, -0.32], $d = .69$; Grade $\times$ Gender: $F(1, 209) = 0.07$, $p = .787$, $\eta^2 = 0.00$, $BF_{\text{excl}} = 4.74$; condition: $F(1, 211) = 0.02$, $p = .976$, $\eta^2 = 0.00$, $BF_{\text{excl}} = 21.74$. These effects indicated that older children had more positive spatial self-concepts than younger children and that girls had lower spatial self-concepts than boys.

Exploratory Analyses of Grade Effects

The grade effects observed for math and spatial attitude measures surprised us, yet we considered the possibility that some of the items on the measures may have been geared toward younger children. For example, older children may encounter less anxiety when performing addition and subtraction problems simply because they have mastered these types of arithmetic operations, whereas younger children, who are still learning these arithmetic

e.g., “taking a math test” or “doing a hard math worksheet”; for details, see [Supplement E in the online supplemental materials](#)) and those that may be more likely to reflect age-specific skills (three items; e.g., “completing a page of addition and subtraction problems”). We then ran repeated measures ANOVAs to examine the factor “math content” (two levels: generic vs. specific) and the effects previously included in analyses of attitudes (i.e., grade and gender). For both math anxiety and math self-concept, there was a

Table 4

Moderating Factors Between Condition and Performance

Model	B (SE)	95% CI	t	p	BF _{incl}	F	df	p	R ²	BF ₁₀
Math anxiety (MA)						7.24	5,138	.,001	.21	2,835.03
Condition (control, math)	2.13 (0.90)		-3.60	.,001	.372	223.81				
MA	-0.76 (0.85)	[[--2.41,		.014		734.05				
Condition × MA	2.90 (1.17)	0.97]4.992.63.,	-0.832.56	.550		12.11				
Gender	0.56 (0.93)	-1.45]0.35]	0.60	.788		0.87				
Grade	0.24 (0.91)		0.27			0.79				
Best Model: Condition, MA, and Condition × M			)							
Math self-concept (MSC)										
Condition	3.08 (0.93)	A (BF[[-1.28, 2.39]1.55, 2.03]10 = 32,515.44	3.32	.001	41.71	4.81	5,132	.,001	.15	33.03
MSC	0.21 (0.84)		0.25	.804	14.83					
Condition × MSC	2.30 (1.09)	[[0.14, 4.45]-1.45,	2.11	.037	5.27					
Gender	0.17 (0.98)	1.87]	0.17	.865	0.76					
Grade	0.47 (0.96)		0.49	.624	0.80					
Best Model: Condition, MSC, a										
Spatial anxiety (SA)										
Spatial self-concept (SSC)		fSC (BF[[-1.77, .10]1.42, 1.87]10 = 23.43)	-0.45	.652	0.08	0.84	5,127	.523	.03	0.02
Best Model: Null Model (none of the included p										
ConditionGradeGenderSACondition (control, space)× SA										
---1.46 (0.89)0.86 (1.19)0.71 (0.90)0.42 (0.83)0.39 (0.86)			-0.730.50	.468	0.04					
redictors; BF[[[[-1.49, 3.22]0.31, 3.22]2.49, 1.06]2.07, 1.23]2.07, 1.64			1.64	.427	0.10					
1.31]= 9.60)			-0.80							
ConditionGenderSSCCondition × SSC--0.17			-0.64	.524	0.08	0.62	5,122	.689	.02	0.03
(1.07)1.02 (0.91)0.05 (0.74)0.55 (0.86)				.950	0.07					
		[[[[-1.95, 2.29]0.79, 2.83]2.65	-0.160.06	.872	0.13					
		0.76]1.51, 1.41]2.25	1.12	.274	0.12					
		1.15] 10.62)	-1.10							
Best Model: Null Model (none of the included										
BFGrade	-0.95 (0.86)									

Note. Significant effects are bolded for ease of viewing. Best models are indicated based on Bayesian model comparisons. We report BF₁₀ or BF_{incl} for all models, regardless of whether it is greater than 1, for ease of presentation. CI = confidence interval; BF = Bayes Factor.

operations, may be more anxious when performing them, an issue highlighted by [Ganley and McGraw \(2016\)](#). This concern is particularly relevant for math attitudes measures and highlights a discrepancy between existing measures in the inclusion of generic versus specific content.

For math attitudes measures, we divided the items into those which were considered generic questions about “math” (five items;

significant Content × Grade interaction such that the simple main effect of grade was only observed for items with specific math content (specific: ps , .001; generic: ps . .54; Content × Grade Interaction: ps , .01, BF₁₀ . 35). That is, when asked about generic math content, anxiety levels did not differ as a function of grade. We consider these findings further in the Discussion section.

For spatial measures, however, this top-down approach was not possible as there was not a clear distinction between general “spatial” items and more specific ones (i.e., each item described a distinct spatial skill). Thus, we ran repeated measures ANOVAs including the new factor “Item” (eight levels) rather than a content factor, as well as grade and gender in the spatial attitudes models. Results revealed grade effects for spatial anxiety for all items except for the item assessing anxiety “identifying which shapes are rectangles and why” (p s , .05). For spatial self-concept, grade effects were observed for only three of the eight items (p s , .05). Full math and spatial models are included in the [online supplemental materials \(Supplement E\)](#).

Exploratory Analyses of Framing Effects

Though we did not observe significant effects of gender on performance following task framing, we considered the possibility that the moderating role of math anxiety and self-concept on performance following math task framing may be stronger for girls, potentially as a result of more negative math attitudes. Thus, we ran exploratory moderation models examining the interaction of Gender $\times$ Math Anxiety and Gender $\times$ Math Self-Concept in the relation between condition and performance. The first model examining the moderating role of math anxiety suggested the effect of gender was not significant, Model: $F(8, 134) = 4.83, p, .001, R^2 = .22, BF_{10} = 117.25$; Condition $\times$ Math Anxiety $\times$ Gender: $B = -0.96, t = -0.79, p = .430, BF_{\text{excl}} = 3.57$, and that the previously observed moderation effect remained significant even when gender effects were included in the model, Condition $\times$ Math Anxiety: $B = 2.11, t = 2.48, p = .015, BF_{\text{incl}} = 1.85$. Further, the model with the most support included only the factors condition, math anxiety, and Condition $\times$ Math Anxiety ($BF_{10} = 16,297.54$). The math self-concept model revealed the same pattern of results, Model: $F(8, 129) = 3.92, p, .001, R^2 = .20, BF_{10} = 32.05$; Condition $\times$ Math Self-Concept $\times$ Gender: $B = 1.78, t = 1.58, p = .116, BF_{\text{incl}} = 3.45$; Condition $\times$ Math Self-Concept: $B = -2.22, t = -2.82, p = .006,$

$BF_{\text{incl}} = 5.36$. Bayesian model comparisons indicated that the best model included the factors Condition, Math Self-Concept, and Condition $\times$ Math Self-Concept. However, we do note that the posterior summary of model terms indicated significant evidence for the interaction between condition, math self-concept, and gender. We interpret these results with caution given the

nonsignificant p value for this term and the fact that this term was not included in the best model of the data.

Generalized Anxiety

Unlike math and spatial anxieties, gender differences for children’s generalized anxiety were not significant, $t(95) = 1.54, p = .128, d = .31, 95\% \text{ CI } [-0.09, 0.71], BF_{01} = 2.40$, suggesting that boys and girls had similar levels of generalized anxiety.

Discussion

The current study found that children’s performance on a novel task varied as a function of how the task was described. Specifically, children were less accurate when a task was described as a math task, relative to when there was no framing. When children heard the task was related to spatial abilities, however, their performance was comparable to performance in the control condition. Notably, moderation analyses revealed that performance differences in our math framing condition varied as a function of differences in math attitudes, not gender. Moreover, despite domain-specific effects of task framing on performance, similar patterns of gender differences in attitudes across math and spatial domains were found. Findings indicated that girls have more negative math and spatial anxieties and self-concepts than boys in early elementary school.

Consequences of Math and Spatial Task Framing

The current work represents the first direct comparison of math and spatial task framing and the impact of this framing on performance. We asked, “can simply highlighting the math or spatial relevance impact a child’s performance?” Based on the current results, we find that when children heard a task was associated with math abilities, they had lower performance relative to children who heard the task was associated with spatial abilities or with no task framing.

Based on prior work, we proposed two hypotheses regarding the relation between framing and performance. The first hypothesis, informed by past work on stereotype threat and lift³ (for related metaanalyses, see Doyle & Voyer, 2016; Flore & Wicherts, 2015), was that gender differences in performance would be observed in the framed conditions. This possibility was predicated on the belief that girls would underperform in the math and spatial framed conditions due to an implicit motivation to align with gender stereotypes, and possibly, that boys may perform better in the framed conditions given that math and space are domains in which their gender is stereotypically expected to excel. Thus, if task framing tapped into children’s beliefs that math and space may be male domains, we would expect to see gender differences in framed conditions.

³ Stereotype threat is the phenomenon that being reminded of one’s membership in a particular group leads to lower performance when that group is stereotypically considered the outgroup for that task (e.g., a female listing her gender on a math test; Shapiro & Williams, 2012). Conversely, stereotype lift is when an individual’s performance is

boosted when they are reminded that an outgroup is negatively stereotyped in the context for which they are about to perform (for review, see Walton & Cohen, 2003).

Our second hypothesis was that math and spatial attitudes may moderate task-framing effects. Under this hypothesis, we predicted that framing a task as related to math or space may tap into individual attitudes toward these domains, such that children with higher domain-specific anxiety (or lower math or spatial self-concept) would have lower performance when the math or spatial relevance of the task was highlighted (relative to no framing).

Results found support for this latter hypothesis. We found that math task framing effects interacted with reported levels of math anxiety and math self-concept. Indeed, the finding that task framing was moderated by anxiety adds to a growing literature on the importance of psychosocial factors (e.g., attitudes, beliefs, motivation) for STEM outcomes. Past work has found that negative math and spatial attitudes are associated with lower performance in children (e.g., Lauer et al., 2018). In the current study, we provide some of the first evidence that these attitudes also can impair performance when it is believed that a task is linked to skills in math. Notably, it was not the case that children with higher math anxiety generally performed worse than children with low math anxiety on our task; importantly, higher math anxiety was not associated with lower performance in the control condition—when the task was not explicitly linked to math ability. Instead, our data suggest it was the child's beliefs about the mathematical nature of the task that likely triggered their math anxiety, with lower performance for children with higher anxiety. Furthermore, similar, though weaker, patterns were found for math self-concept suggesting that multiple math attitudes may be activated when hearing that a task is linked to mathematical abilities. These findings imply that math anxiety and math self-concept have the potential to play a salient role in children's learning and achievement regardless of the nature of the actual task at hand. Critically, we observed these findings in children who are within the first years of their educational career highlighting an early-emerging mechanism that may impede later STEM pursuits and achievement.

Notably, analyses did not reveal gender differences in response to framing effects. As such, it seems unlikely that stereotypes played a major role in our pattern of results. Why is this the case? One possibility is that school children may be less likely to experience stereotype threat in math or spatial contexts than their adult counterparts. Evidence of stereotype threat in math contexts with children is mixed (for metaanalyses, see Doyle & Voyer, 2016; Flore & Wicherts, 2015), and as such, it has been hypothesized that stereotype threat may not reliably be observed until later in development. Moreover, there is no evidence of spatial stereotype threat in childhood (Doyle & Voyer, 2016), consistent with our findings. However, it should be pointed out that our paradigm was not specifically designed to elicit stereotype threat in children—our manipulation was more subtle than in studies of stereotype threat and in fact, involved no gender activation—a factor important for tapping gender stereotypes. Additionally, though there is evidence that children hold gender stereotypes associating boys with math and spatial domains, findings are mixed (studies supporting gender stereotypes in childhood: e.g., Cvencek et al., 2011; Neuburger et al., 2015; studies finding no evidence or

evidence of own-gender biases: Martinot et al., 2012; Moè, 2018). Given that we did not include any assessment of whether the children in our study hold math or spatial gender stereotypes, we are unable to ascertain whether children in our sample were even susceptible to stereotype threat. Therefore, the lack of gender effects may be attributable to developmental patterns, the actual manipulation, or both. Adjudicating between these mechanisms in future studies will be critical for understanding the factors that impact girls and women in STEM settings.

Domain Differences

Despite observing many similarities between math and spatial domains regarding gender differences in domain-specific attitudes, lower performance following task framing was uniquely observed for math. In fact, Bayesian analyses suggested that there was evidence against performance differences in the spatial task framing condition (as compared to the control condition).

Why was math framing linked to performance but spatial framing was not? One possibility is that differences in children's experiences with math versus space may account for these results. Specifically, children's early spatial experiences primarily consist of informal, playful activities that require spatial reasoning (e.g., puzzles, building blocks; e.g., Dearing & Tang, 2010; Huntsinger et al., 2016; Jirout & Newcombe, 2015). This is in indirect contrast to early math experiences, much of which occurs in formal settings (i.e., school) and involve explicit instruction, feedback, and evaluation. In turn, this divergence may lead to distinct patterns in children's motivation, representation, and performance in math versus spatial contexts. For example, the formal setting of early math experiences presents frequent opportunities for children to have negative experiences with math such as difficulty learning or receiving negative feedback and evaluation. These experiences can have a profound impact on the development of math anxiety and motivation on math tasks, both of which are suggested to negatively impact performance (Maloney & Beilock, 2012). Thus, simply highlighting the mathematical nature of a task may be sufficient to activate negative math attitudes, in part because math experiences are more salient. Conversely, though children may also experience difficulty or negative feedback during spatial play activities, these negative experiences are likely infrequent and less salient when they do occur as they are not a typical feature of informal play.

Children may also feel more pressure to do well on a math task than on a spatial task. Specifically, it has been theorized that performance pressure—which is associated with lower performance (Baumeister, 1984)—is more salient for tasks that are viewed as important (Baumeister, 1984), and past work has found that young children believe that math is both useful and important (e.g., Leedy et al., 2003). However, though no empirical work has examined these types of beliefs in the domain of space, it would be surprising if similar patterns were observed. That is, the importance of spatial reasoning may be much more abstract in its perceived importance than math which is explicitly linked to achievement. If children view math skills as more important than spatial skills, they would likely feel more pressure to perform on a math task, and thus, be more susceptible to performance impacts. With these differences in

performance pressures, it may be the case that math attitudes play a greater role in math performance than spatial attitudes play in spatial performance. Future work should examine whether performance is impacted following spatial framing as the importance of spatial skills becomes less abstract (i.e., at older ages). Further, future work should explicitly examine the perceived importance of math and space as a potential moderator of observed performance.

Importantly, results cannot be attributed to a more general pattern linking performance to task framing. That is, past work has suggested that generic statements and essentialist language (e.g., smart people are good at math) predict decrements in performance, especially for individuals with negative attitudes toward the domain highlighted (e.g., Cimpian, 2010; Cimpian et al., 2007, 2012). This has even been found for novel tasks such that when performance on the task was linked to a social category (e.g., “boys”), children’s performance suffered, regardless of whether or not they actually belonged to the social group (Cimpian et al., 2012). However, if essentialist beliefs drive framing effects, we would expect a link between lower performance and task framing in both the math and spatial conditions. Nevertheless, future work should consider the role that essentialist beliefs and motivational frameworks may play in performance in math and spatial contexts.

Early-Emerging Gender Differences in Math and Spatial Attitudes

The current study is the first to characterize multiple math and spatial attitudes, specifically anxieties and self-concepts, in elementary school-aged children and to assess gender differences in these attitudes in children experiencing the first year of formal schooling. Overall, we found moderately sized gender differences in children’s math and spatial anxieties and self-concepts, which were similar between younger and older children. Further, per Bayesian analyses, the evidence for gender differences was very strong or decisive for all attitudes suggesting strong support for early-emerging gender differences in children’s math and spatial anxieties and self-concepts. These findings provide key insights into the children’s early math and spatial attitudes.

Math Attitudes

In light of conflicting findings on the existence of gender differences in math anxiety across children in early elementary school (e.g., Ahmed et al., 2012; Harari et al., 2013; Jameson, 2014; Lauer et al., 2018), we found a significant overall gender difference in children’s math anxiety. Specifically, we found very strong evidence (as indicated by $BF_{10} = 73.44$) that girls indicated greater math anxiety than boys, in line with some previous work (e.g., Lauer et al., 2018). Gender differences in children’s math self-concepts similarly indicated that girls rated their own math abilities more poorly than their male peers, consistent with previous work (Ahmed et al., 2012; Cvencek et al., 2011, 2015). For both math anxiety and self-concept, we did not observe significant interactions between gender and grade. Thus, our current results suggest that gender differences in math anxiety and self-concept are already

evident in children as young as six and that these gender differences are relatively stable regardless of the child’s grade.

Spatial Attitudes

In line with limited previous work (e.g., Lauer et al., 2018; Ramirez et al., 2012), girls reported significantly higher spatial anxiety than boys, with an overall moderate effect of gender with decisive support for the effect ($d = 0.56$, $BF_{10} = 470.60$). Critically, given that gender differences in spatial anxiety are consistently documented in adulthood (e.g., Alvarez-Vargas et al., 2020; Sokolowski et al., 2019), the present work extends these findings to reveal that these gender differences are observable and are stable (i.e., no group differences across children in different grades) fairly early in development.

Further, we provided the first examination of children’s spatial self-concept. Again, findings were consistent with prior work on spatial self-concept in adults (e.g., Blajenkova et al., 2006; Hegarty et al., 2002), with young girls indicating lower spatial self-concept than boys. Together, our findings suggest that, as early as first grade, girls hold significantly more negative attitudes toward STEM-related domains than boys.

Understanding Gender Differences in Children’s Math and Spatial Attitudes

Why are gender differences in math and spatial anxiety and self-concept already present within the first years of experience with formal education? Some have proposed that early-emerging gender differences in anxiety, in particular, may simply reflect general patterns of gender differences in anxiety, regardless of the type of anxiety (for review, see Dowker et al., 2016). Though previous work has noted that girls are more likely to report anxiety (e.g., Dowker et al., 2016), and have higher rates of anxiety disorders (e.g., McLean et al., 2011), this was not the case with children in our sample. Specifically, we did not observe gender differences in children’s generalized anxiety. Thus, it is unlikely that gender differences in math and spatial anxiety are a simple byproduct of gender differences in generalized anxiety.

Additionally, it is unlikely that observed gender differences in math and spatial attitudes can be attributed to gender differences in math and spatial skills. Although there is consistent evidence of gender differences in some spatial skills, such as mental rotation (for recent meta-analysis, see Lauer et al., 2018), reports of gender differences in math skills are uncommon (Hutchison et al., 2019). As such, if gender differences in math and spatial attitudes are perpetuated by gender differences in performance, then we would only expect to observe gender differences in spatial attitudes. Given that boys and girls perform comparably on math tasks, then there would be no reason for girls to have higher math anxiety than boys—yet this is the case.

Instead, perceptions of gender differences of who is associated with math and spatial domains, rather than actual differences in skills, provide the most parsimonious account of gender differences in math and spatial attitudes. We suggest that gender differences in math and spatial attitudes may be an early-emerging consequence

of pervasive gender stereotypes that math and space are “male” domains. Some prior work has found that children already associate males with both math and spatial domains as early as elementary school (Cvencek et al., 2011; Neuburger et al., 2015), suggesting that gender differences in math and spatial self-concept exist at the same period of development in which children’s math and spatial gender stereotypes reflect cultural gender stereotypes. Further, gender differences in math and spatial self-concepts had more decisive support (per Bayesian analyses), and were largereffects, than gender differences in math and spatial anxieties. This is relevant because anxiety is posited to be less proximally related to one’s identity than their self-concept, suggesting that math and spatial selfconcepts may provide the foundation for math and spatial anxiety.

Grade Effects for Math and Spatial Attitudes

One particularly interesting finding from the current work was that older children indicated more positive math and spatial attitudes (lower anxiety and higher self-concepts) than their younger counterparts. Notably, this finding conflicts with theories suggesting relatively stable math attitudes across early childhood (Wang et al., 2020), as well as the alternative theory that formal math experiences lead to increased negative math attitudes (e.g., Krinzinger et al., 2009). However, this finding should be interpreted with caution. Both the math and spatial attitudes measures were adapted from measures designed to be used with first and second graders. Though we made effortsto adapt itemsto be developmentallyappropriate for third and fourth graders, it is possible that age effects were an artifact of the specific content of the measure and its intended target population. Indeed, the content may have simply been viewed as easier and thus, older children reported lower anxiety and higher self-concept. Post hoc exploratory examinations of grade effects on specific versus generic items supported this possibility for both math anxiety and math self-concept measures. Specifically, the effect of grade was significant for items that included age-specific math content (e.g., completing a page of addition and subtraction problems), but not on items about generic math experiences (e.g., “taking a math test”). Our findings align with the only known previous study to address concerns regarding age effects and the content of children’s math anxiety questionnaires (Ganley & McGraw, 2016). Indeed, that work found significant age effects only for the specific math anxiety item in their questionnaire. Futurework should explore these age effects by using more general math attitude measures that do not refer to specific math abilities (e.g., arithmetic) to explore whether these trends hold using a less age-specific measure.

Post hoc examinations of grade effects on measures of spatial attitudes were less clear. Specifically, for spatial anxiety, grade effects were significant for all but one item and for three of eight spatial selfconcept items. Why might age effects be more prominent for spatial attitudes than math attitudes? One possibility is that children’s early spatial experiences are generally playful (block play, Legos, etc.). Thus, children’s spatial attitudes may become more positive with age because spatial activities have low stakes. A similar explanation in the math domain, however, does not seem as

likely. Nevertheless, it is possible that spatial attitudes may become more negative as spatial content becomes more prominent in math and science classes later in schooling (e.g., in high school, when students encounter physics). If so, a developmental divergence between math and spatial attitudes may emerge due to distinct experiences with math and spatial content across development. Future research may consider exploring math and spatial attitudes using a longitudinal design to explore how these attitudes change over the course of early to middle childhood.

Implications for Understandings of Math–Space Relations

The current work provides some of the first evidence for robust similarities between math and spatial domains, specifically, that these similarities are observed for attitudes in addition to skills. There is strong theoretical and empirical support for relations between math and spatial skills and achievement (e.g., Gunderson & Hildebrand, 2021; Mix & Cheng, 2012; Newcombe et al., 2015). In fact, some theorists have even argued that mathematical and spatial processing arise from the same magnitude processing system early in development, though others posit that quantity representations are better represented as similar but distinct systems (for review, see Hamamouche & Cordes, 2019). Recent empirical evidence supports the latter conceptualization finding that math and spatial domains are highly related but distinct (e.g., Mix et al., 2016, 2017). Critically, only recently have researchers suggested that the mathspace relation may include processes beyond skills and achievement (e.g., Lauer et al., 2018; Sokolowski et al., 2019). The current work aligns with theoretical accounts suggesting math and spatial processes are similar but distinct, and suggests these processes include attitudes in addition to skills. Specifically, we observed multiple similarities in children’s math and spatial attitudes, but a divergence in math and spatial framing effects, as discussed above.

One of the most striking similarities between math and spatial attitudes pertains to the consistent gender differences across attitudes. However, it is possible that gender differences in math and spatial anxieties and self-concepts are not unique to these domains and instead reflect a distinction between girls’ and boys’ attitudes more generally across early childhood. Critically, however, we found no gender differences in children’s generalized anxiety, a finding that may be taken as evidence that this finding is unique to the math-space relation. Future work should examine the relation between domain-specific and generalized attitudes more systematically to provide a full picture of how gender differences in math and spatial attitudes emerge in childhood. For example, studying the development of these attitudes across early childhood (i.e., longitudinally), rather than characterizing age-related differences (i.e., cross-sectionally as in the current study), is essential for understanding when and why these gender differences emerge. Regardless, our results highlight elementary school as a critical point in development for gender differences in math and spatial attitudes that occurs prior to gender differences in STEM-relevant outcomes (e.g., STEM participation).

Limitations

A key limitation to consider when interpreting the current work is our choice of task for assessing performance under specific task framing. The task, nonsymbolic numerical magnitude comparison, is classically defined as a numerical task. For this reason, the finding that performance was only different from baseline in the math condition might be explained by factors related to the believability that the task was spatial. However, we do not believe this to be the case for a few reasons. First, spatial skills are highly related to nonsymbolic numerical comparison performance (Carr et al., 2020), and performance is impacted by spatial aspects of the task (i.e., size and arrangement of dots; DeWind et al., 2015). Further, we adapted the directions of the task to ensure we did not use any numerical language that may serve as a cue to the mathematical nature of the task. Moreover, the task was likely novel to most children and did not include symbolic representations of number (i.e., Arabic numerals) making it unlikely that children would automatically assume that the task is a math task. Finally, in a nearly identical study conducted with adults (Hildebrand et al., under review), both math and spatial framing were linked to adults' performance on the nonsymbolic numerical comparison task as well as on a more spatial task (matrix reasoning). Because adults' performance was related to spatial framing on this very same task, this may suggest developmental differences in the salience of spatial framing, rather than an inability to interpret the task as spatial in nature.

Further, it is an open question whether differences in performance across our groups may have been observed at pretest. It should be noted that we used a full randomized controlled trial design, in which children were randomly assigned to conditions at recruitment. Moreover, children in our three conditions did not differ significantly in terms of where they were recruited (school, lab, museums, etc.) or in terms of their general demographics, and notably, we found no group differences in any of our additional measures (math and spatial attitudes, working memory, generalized anxiety; p s > .30). Thus, we can reasonably conclude that performance differences across conditions are attributed to our task framing manipulation and not another confounding factor. However, future research may wish to include a pretest measure to allow for a stronger exploration of the impact of task framing on performance within the same individuals.

An additional limitation stems from the likelihood that the concept of "spatial skills" is less familiar to children than the concept of "math skills." To mitigate this concern, in both the math and spatial conditions, we made sure to provide thorough introductions to the concept of math (or spatial) skills. These introductions included information about the specific skills involved in these domains, examples of tasks in these domains, and ensuring the child could provide an additional valid example. It is also important to note that we observed similar age-related patterns of gender differences in attitudes across domains. Further, Bayesian analyses revealed stronger support for gender differences for each spatial attitude than for the matched math attitude. Thus, it appears that even if the concept of "spatial skills" is unfamiliar to children, our results found robust gender differences suggesting that children did understand the concepts. However, developing a more

comprehensive understanding of children's conceptual understanding and representations of math and space could provide key insights for basic and applied work related to math and spatial processes.

We also note key considerations regarding the generalizability of the work at hand. First, our most complex analyses were powered only to assess grade differences across two-grade bins. Thus, it would be helpful for future work to include larger samples in order to look at age effects within specific grades. Additionally, our sample was recruited and tested in various settings. As such, it is possible there were additional factors that impacted results.

Conclusion

In sum, the current work demonstrates that framing a task as a math task leads to poorer performance, especially for children who are higher in math anxiety and lower in math self-concept. However, this pattern was not found for spatial framing. In addition, our findings revealed that robust gender differences in math and spatial attitudes are present early in childhood. Together, our results suggest that children's performance in math may be impacted before it is even impacted by spatial framing. That is, future work may build upon these findings to examine whether parents or teachers may inadvertently tap into children's negative math attitudes simply by talking about what type of task they are about to perform, which in turn, may represent a powerful, early factor that inhibits certain individuals' STEM participation and success. Fortunately, mitigating these effects may be as simple and straightforward as changing the way we talk about tasks. Findings can be leveraged to advance our understanding of influences on STEM outcomes that emerge early in development.

References

- Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences, 22*(3), 385–389. <https://doi.org/10.1016/j.lindif.2011.12.004>
- Alvarez-Vargas, D., Abad, C., & Pruden, S. M. (2020). Spatial anxiety mediates the sex difference in adult mental rotation test performance. *Cognitive Research: Principles and Implications, 5*(1), Article 31. <https://doi.org/10.1186/s41235-020-00231-8>
- Arens, A. K., Frenzel, A. C., & Goetz, T. (2022). Self-concept and self-efficacy in math: Longitudinal interrelations and reciprocal linkages with achievement. *The Journal of Experimental Education, 90*(3), 615–633. <https://doi.org/10.1080/00220973.2020.1786347>
- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181–185. <https://doi.org/10.1111/1467-8721.00196>
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*(2), 224–237. <https://doi.org/10.1037/0096-3445.130.2.224>
- Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243–248. <https://doi.org/10.3758/BF03194059>

- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, 147(2), 134–168. <https://doi.org/10.1037/bul0000307>
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46(3), 610–620. <https://doi.org/10.1037/0022-3514.46.3.610>
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(4), 545–553. <https://doi.org/10.1097/00004583-199704000-00018>
- Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 20(2), 239–263. <https://doi.org/10.1002/acp.1182>
- Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, 34(3), 139–153. https://doi.org/10.1207/s15326985ep3403_1
- Carr, M., Horan, E., Alexeev, N., Bamed, N., Wang, L., & Otumfuor, B. (2020). A longitudinal study of spatial skills and number sense development in elementary school children. *Journal of Educational Psychology*, 112(1), 53–69. <https://doi.org/10.1037/edu0000363>
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135(2), 218–261. <https://doi.org/10.1037/a0014412>
- Cimpian, A. (2010). The impact of generic language about ability on children's achievement motivation. *Developmental Psychology*, 46(5), 1333–1340. <https://doi.org/10.1037/a0019665>
- Cimpian, A., Arce, H.-M. C., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychological Science*, 18(4), 314–316. <https://doi.org/10.1111/j.1467-9280.2007.01896.x>
- Cimpian, A., Mu, Y., & Erickson, L. C. (2012). Who is good at this game? Linking an activity to a social category undermines children's achievement. *Psychological Science*, 23(5), 533–541. <https://doi.org/10.1177/0956797611429803>
- Cvencek, D., Kapur, M., & Meltzoff, A. N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning and Instruction*, 39(Supplement C), 1–10. <https://doi.org/10.1016/j.learninstruc.2015.04.002>
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children. *Child Development*, 82(3), 766–779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>
- Daker, R. J., Delage, V., Maloney, E. A., & Lyons, I. M. (2022). Testing the specificity of links between anxiety and performance within mathematics and spatial reasoning. *Annals of the New York Academy of Sciences*, 1512(1), 174–191. <https://doi.org/10.1111/nyas.14761>
- Dearing, E., & Tang, S. (2010). The home learning environment and achievement during childhood. In S. Christenson & A. L. Reschly (Eds.), *Handbook on school-family partnerships* (pp. 131–157). Taylor & Francis.
- Delage, V., Trudel, G., Retanal, F., & Maloney, E. A. (2021). Spatial anxiety and spatial ability: Mediators of gender differences in math anxiety. *Journal of Experimental Psychology: General*, 151(4), 921–933. <https://doi.org/10.1037/xge0000884>
- DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, 142, 247–265. <https://doi.org/10.1016/j.cognition.2015.05.016>
- Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, 25(1), 207–218. <https://doi.org/10.3758/s13423-017-1266-z>
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, 7, Article 508. <https://doi.org/10.3389/fpsyg.2016.00508>
- Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences*, 47, 103–116. <https://doi.org/10.1016/j.lindif.2015.12.018>
- Dweck, C. S. (2007a). Is math a gift? Beliefs that put females at risk. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science?: Top researchers debate the evidence* (pp. 47–55). American Psychological Association. <https://doi.org/10.1037/11546-004>
- Dweck, C. S. (2007b). The secret to raising smart kids. *Scientific American Mind*, 18(6), 36–43. <https://doi.org/10.1038/scientificamericanmind1207-36>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53(1), 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>
- Ganley, C. M., & McGraw, A. L. (2016). The development and validation of a revised version of the math anxiety scale for young children. *Frontiers in Psychology*, 7, Article 1181. <https://doi.org/10.3389/fpsyg.2016.01181>
- Gunderson, E. A., & Hildebrand, L. (2021). Relations among spatial skills, number line estimation, and exact and approximate calculation in young children. *Journal of Experimental Child Psychology*, 212, Article 105251. <https://doi.org/10.1016/j.jecp.2021.105251>
- Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2017). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development*, 19(1), 21–46. <https://doi.org/10.1080/15248372.2017.1421538>
- 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(7213), 665–668. <https://doi.org/10.1038/nature07246>
- Hamamouche, K., & Cordes, S. (2019). Number, time, and space are not singularly represented: Evidence against a common magnitude system beyond early childhood. *Psychonomic Bulletin & Review*, 26(3), 833–854. <https://doi.org/10.3758/s13423-018-1561-3>
- Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. *The Journal of Experimental*

- Education, 81(4), 538–555. <https://doi.org/10.1080/00220973.2012.727888>
- Hart, S. A., & Ganley, C. M. (2019). The nature of math anxiety in adults: Prevalence and correlates. *Journal of Numerical Cognition*, 5(2), 122–139. <https://doi.org/10.5964/jnc.v5i2.195>
- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53(1), 87–97. <https://doi.org/10.2307/1129640>
- Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30(5), 425–447. [https://doi.org/10.1016/S0160-2896\(02\)00116-2](https://doi.org/10.1016/S0160-2896(02)00116-2)
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21(1), 33–46. <https://doi.org/10.2307/749455>
- Hildebrand, L., & Cordes, S. (2023). Performance impacts of STEM labels. <https://doi.org/10.17605/OSF.IO/Y9BMH>
- Hildebrand, L., Posid, T., Hymes, L., Moss-Racusin, C. A., & Cordes, S. (2022). Does my daughter like math? Relations between parent and child math attitudes and beliefs. *Developmental Science*, 26(1), Article e13243. <https://doi.org/10.1111/desc.13243>
- Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, 48, 45–53. <https://doi.org/10.1016/j.lindif.2016.02.006>
- Hirnsstein, M., Coloma Andrews, L., & Hausmann, M. (2014). Gender stereotyping and cognitive sex differences in mixed- and same-sex groups. *Archives of Sexual Behavior*, 43(8), 1663–1673. <https://doi.org/10.1007/s10508-014-0311-5>
- Huntsinger, C. S., Jose, P. E., & Luo, Z. (2016). Parental facilitation of early mathematics and reading skills and knowledge through encouragement of home-based activities. *Early Childhood Research Quarterly*, 37, 1–15. <https://doi.org/10.1016/j.ecresq.2016.02.005>
- 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>
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A metaanalysis. *Psychology of Women Quarterly*, 14(3), 299–324. <https://doi.org/10.1111/j.1471-6402.1990.tb00022.x>
- Jameson, M. M. (2014). Contextual factors related to math anxiety in secondgrade children. *The Journal of Experimental Education*, 82(4), 518–536. <https://doi.org/10.1080/00220973.2013.813367>
- JASP Team. (2021). JASP (Version 0.16) [Computer software]. <https://jaspstats.org/>
- Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science*, 26(3), 302–310. <https://doi.org/10.1177/0956797614563338>
- Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, 73(1), 1–2. <https://doi.org/10.1037/amp0000263>
- Kriegbaum, K., Jansen, M., & Spinath, B. (2015). Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement. *Learning and Individual Differences*, 43, 140–148. <https://doi.org/10.1016/j.lindif.2015.08.026>
- Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeducational Assessment*, 27(3), 206–225. <https://doi.org/10.1177/0734282908330583>
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Lauer, J. E., Esposito, A. G., & Bauer, P. J. (2018). Domain-specific anxiety relates to children's math and spatial performance. *Developmental Psychology*, 54(11), 2126–2138. <https://doi.org/10.1037/dev0000605>
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19(3), 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- Leedy, M. G., LaLonde, D., & Runk, K. (2003). Gender equity in mathematics: Beliefs of students, parents, and teachers. *School Science and Mathematics*, 103(6), 285–292. <https://doi.org/10.1111/j.1949-8594.2003.tb18151.x>
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, 14(6), 1292–1300. <https://doi.org/10.1111/j.1467-7687.2011.01080.x>
- Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition*, 4(3), 526–553. <https://doi.org/10.5964/jnc.v4i3.154>
- Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, 16(8), 404–406. <https://doi.org/10.1016/j.tics.2012.06.008>
- Martinot, D., Bagès, C., & Désert, M. (2012). French Children's awareness of gender stereotypes about mathematics and reading: When girls improve their reputation in math. *Sex Roles*, 66(3–4), 210–219. <https://doi.org/10.1007/s11199-011-0032-3>
- Master, A. (2021). Gender stereotypes influence children's stem motivation. *Child Development Perspectives*, 15(3), 203–210. <https://doi.org/10.1111/cdep.12424>
- McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, 45(8), 1027–1035. <https://doi.org/10.1016/j.jpsychires.2011.03.006>
- Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior*, 42, 197–243. <https://doi.org/10.1016/B978-0-12-394388-0.00006-X>
- 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>
- 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*, 18(4), 465–492. <https://doi.org/10.1080/15248372.2017.1346658>
- Moè, A. (2018). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and*

- Individual Differences, 61, 172–180. <https://doi.org/10.1016/j.lindif.2017.12.002>
- National Center for Science and Engineering Statistics. (2021). Women, Minorities, and Persons with Disabilities in Science and Engineering: 2021 (Special Report NSF). National Science Foundation. <https://nces.gov/wmpd>
- Neuburger, S., Ruthsatz, V., Jansen, P., & Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learning and Individual Differences*, 37, 169–175. <https://doi.org/10.1016/j.lindif.2014.09.003>
- Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(6), 491–505. <https://doi.org/10.1002/wcs.1369>
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math = me. *Journal of Personality and Social Psychology*, 83(1), 44–59. <https://doi.org/10.1037/0022-3514.83.1.44>
- Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2012). Spatial anxiety relates to spatial abilities as a function of working memory in children. *The Quarterly Journal of Experimental Psychology*, 65(3), 474–487. <https://doi.org/10.1080/17470218.2011.616214>
- Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, 14(2), 187–202. <https://doi.org/10.1080/15248372.2012.664593>
- R Core Team. (2020). R: 2019. A language and environment for statistical computing version, 3(1).
- Rebok, G., Riley, A., Forrest, C., Starfield, B., Green, B., Robertson, J., & Tambor, E. (2001). Elementary school-aged children's reports of their health: A cognitive interviewing study. *Quality of Life Research*, 10(1), 59–70. <https://doi.org/10.1023/A:1016693417166>
- Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, 66(3-4), 175–183. <https://doi.org/10.1007/s11199-011-0051-0>
- Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition*, 182, 193–212. <https://doi.org/10.1016/j.cognition.2018.10.005>
- Tokita, M., & Ishiguchi, A. (2013). Effects of perceptual variables on numerosity comparison in 5–6-year-olds and adults. *Frontiers in Psychology*, 4, Article 431. <https://doi.org/10.3389/fpsyg.2013.00431>
- Vander Heyden, K. M., van Atteveldt, N. M., Huizinga, M., & Jolles, J. (2016). Implicit and explicit gender beliefs in spatial ability: Stronger stereotyping in boys than girls. *Frontiers in Psychology*, 7, Article 1114. <https://doi.org/10.3389/fpsyg.2016.01114>
- Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology*, 38(1), 1–10. <https://doi.org/10.1016/j.cedpsych.2012.09.001>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Droppedmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., ... Morey, R. D. (2018). Bayesian Inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39(5), 456–467. [https://doi.org/10.1016/S0022-1031\(03\)00019-2](https://doi.org/10.1016/S0022-1031(03)00019-2)
- Wang, M.-T., Degol, J., & Ye, F. (2015). Math achievement is important, but task values are critical, too: Examining the intellectual and motivational factors leading to gender disparities in STEM careers. *Frontiers in Psychology*, 6, 36. <https://doi.org/10.3389/fpsyg.2015.00036>
- Wang, M.-T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, 24(5), 770–775. <https://doi.org/10.1177/0956797612458937>
- Wang, Z., Oh, W., Malanchini, M., & Borriello, G. A. (2020). The developmental trajectories of mathematics anxiety: Cognitive, personality, and environmental correlates. *Contemporary Educational Psychology*, 61, Article 101876. <https://doi.org/10.1016/j.cedpsych.2020.101876>
- Wechsler, D. (2003). Wechsler intelligence scale for children—Fourth Edition (WISC-IV). The Psychological Corporation.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A grammar of data manipulation* (Version 1.0.5). <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>
- Wong, W. I. (2017). The space-math link in preschool boys and girls: Importance of mental transformation, targeting accuracy, and spatial anxiety. *British Journal of Developmental Psychology*, 35(2), 249–266. <https://doi.org/10.1111/bjdp.12161>
- Yang, J., Rahardja, S., & Fränti, P. (2019). Outlier detection: How to threshold outlier scores? *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing* (pp. 1–6). <https://doi.org/10.1145/3371425.3371427>

Received February 9, 2022

Revision received June 9, 2023

Accepted June 23, 2023 ■