Journal of Experimental Child Psychology 211 (2021) 105222

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Boys' visuospatial abilities compensate for their relatively poor in-class attentive behavior in learning mathematics

David C. Geary *, Mary K. Hoard, Lara Nugent

Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA

A R T I C L E   I N F O

Article history:
Received 6 January 2021
Revised 18 May 2021
Accepted 5 June 2021

Keywords:
Mathematics achievement
In-class attention
Sex differences
Spatial abilities
Attention deficit disorder
Hyperactivity

A B S T R A C T

The mathematics and reading achievement of 322 adolescents (159 boys) was assessed in seventh and eighth grades, as were their intelligence, working memory, and spatial abilities. Their seventh- and eighth-grade mathematics and English language arts teachers reported on their in-class attentive behavior. The latter emerged as an important predictor of achievement, but more so for mathematics than for reading. Boy were less attentive in classroom settings than girls (d = −.34) and performed better than expected in mathematics given their level of engagement in the classroom. Boys' better-than-expected mathematics achievement was related to advantages on visuospatial measures (ds = .28–.56), which fully mediated a sex difference in mathematics (ds = .27–.28) but not in reading achievement, with control of in-class attentive behavior. The results suggest that boys' advantages in visuospatial skills compensate for lower levels of classroom engagement in the learning of mathematics but not in reading competencies.

* Corresponding author.
E-mail address: gearyd@missouri.edu (D.C. Geary).

https://doi.org/10.1016/j.jecp.2021.105222

Introduction

Lower attentional and behavioral engagement in classroom settings is consistently related to lower academic achievement, controlling for intelligence, parental socioeconomic status, and prior achievement (Duncan et al., 2007; Polderman, Boomsma, Bartels, Verhulst, & Huizink, 2010). A large-scale observational study of mathematics classrooms revealed that children who were frequently engaged in off-task behaviors (e.g., talking to friends) had below average mathematics achievement ($r = -.47$), whereas their more attentive (e.g., focused on teacher) peers had above average achievement ($r = .37$) (Stigler, Lee, & Stevenson, 1987). Lower achievement in turn is associated with lower occupational and socioeconomic attainment during adulthood (Rivera-Batiz, 1992; Ritchie & Bates, 2013). Poor engagement in classroom settings thus has potentially lifelong consequences, controlling for other factors.

Among children and adolescents with attention-deficit/hyperactivity disorder (ADHD), there are at least two boys to every girl (Ramtekkar, Reiersen, Todorov, & Todd, 2010). For typically developing children, boys consistently score lower than girls on more continuous measurements of attentive behavior (Rietveld, Hudziak, Bartels, Van Beijsterveldt, & Boomsma, 2004; Willcutt, 2012). Given these relations, we might expect that boys, on average, will show lower academic achievement than girls. This is indeed the case for reading (Reilly, Neumann, & Andrews, 2019), which appears to be partially related to the sex difference in attentive behavior (Daucourt, Erbeli, Little, Haughbrook, & Hart, 2020; Willcutt & Pennington, 2000), and compromises their long-term educational prospects (Stoet & Geary, 2020).

At the same time, boys score as well as, and often better than, girls on mathematics achievement tests (Hyde, Fennema, & Lamon, 1990; Lindberg, Hyde, Petersen, & Linn, 2010; Stoet & Geary, 2013) despite less attentive behavior in classroom settings. One possibility is that the sex difference on these tests is driven by more boys than girls at the high end of performance (Wai, Hodges, & Makel, 2018). However, this is not the whole story. In a large-scale multinational study of nearly 1.5 million adolescents, Stoet & Geary, 2013 showed that the performance of boys was at least as good as, and typically better than, that of girls across the entire continuum of mathematics achievement. There must be some mechanism that supports the learning of mathematics and that compensates for boys' lower levels of classroom engagement. The current study tested the hypothesis that this mechanism is visuospatial abilities (Casey, Nuttall, & Pezaris, 1997; Geary, 1996; Halpern et al., 2007).

Sex differences in attentive behavior

Teacher and parent ratings of children's distractibility, attention to detail, organization, and so forth are consistently related to behavior in school (e.g., staying seated) and to concurrent achievement and longitudinal gains in achievement (Duncan et al., 2007; Fuchs et al., 2006; Fuchs, Geary, Compton, Fuchs, & Hamlett, 2014; Geary, Hoard, Nugent, & Bailey, 2013; Gray, Dueck, Rogers, & Tannock, 2017; Polderman, Boomsma, Bartels, Verhulst, & Huizink, 2010). In classroom settings, these behaviors manifest as sustained attention and attention to details during school-related activities as well as distractibility during these activities (Swanson et al., 2012). Teacher ratings of attentional control in theory are related to conceptually similar cognitive abilities, such as working memory and executive functions, but teacher ratings are predictive of achievement outcomes above and beyond these abilities. In other words, the ability to maintain attentional focus on academic material in actual classrooms is not fully captured by standard tests of cognitive abilities (Friso-van den Bos & van de Weijer-Bergsma, 2020).

As noted, boys are overrepresented among children and adolescents who have been diagnosed with ADHD (Willcutt, 2012), and they are rated by parents and teachers as less attentive and more hyperactive than same-age girls (Arnett, Pennington, Willcutt, DeFries, & Olson, 2015). As children move from the preschool years into elementary school and adolescence, hyperactivity decreases but the sex difference in inattention remains and is the primary correlate with achievement. Using continuous measures in a longitudinal study that included more than 10,000 twins, Rietveld et al. (2004) found moderately lower attentive behavior in boys than in girls during elementary school (ds = $-.35$ to $-.38$) and adolescence ($d = -.37$) (see also Lampert, Polanczyk, Tramontina, Mardini, & Rohde, 2004). Inattention had similar heritable influences (>70% of variance) on boys and girls, and boys and girls showed similar levels of longitudinal stability after the preschool years, although there may be an increased risk of late onset hyperactivity (but not inattention) for some adolescent girls (Murray et al., 2019). Whatever the developmental trajectories, parent and teacher ratings of attentive behavior are equally predictive of achievement (external validity) for girls and boys (Arnett et al., 2015), indicating that there is not a bias to rate boys as less attentive than they actually are.

Visuospatial abilities and mathematics achievement

The overall sex difference in mathematics achievement generally favors boys but is small in magnitude ($d \sim .10$) and can vary across topics, grades, and countries (Hyde, Fennema, & Lamon, 1990; Lindberg, Hyde, Petersen, & Linn, 2010; Stoet & Geary, 2013). Sex differences are often found at the high end of performance (Wai et al., 2018), for more advanced topics (Hyde et al., 1990), and in areas that have a spatial component to them (Geary, 1996; Halpern et al., 2007), but the reasons for these differences are debated (Ceci & Williams, 2010; Hawes & Ansari, 2020; Mix, 2019). Whatever the reasons, a relation between visuospatial abilities and mathematics achievement in school and in math-intensive occupations has been extensively documented (Hawes & Ansari, 2020; Kell, Lubinski, Benbow, & Steiger, 2013; Mix, 2019). In fact, a relation between early-developing spatial skills (e.g., noticing change in spatial orientation) and later performance in arithmetic is found during the preschool years (Lauer & Lourenco, 2016; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017) and early elementary school years (Cheng & Mix, 2014; Cheung, Sung, & Lourenco, 2019).

As an example, in two large-scale studies, Mix and colleagues (2016, 2017) administered a battery of mathematics and visuospatial measures to kindergarteners and third- and sixth-graders. In each grade, the tests formed distinct but highly correlated mathematics and visuospatial factors (rs = .55–.73); the pattern remained when vocabulary, an index of intelligence, was controlled. The relations between the two factors were stronger than the relations between any specific mathematics or visuospatial measures, suggesting that the relation extends across mathematics and spatial tests. There were also some indications that visuospatial abilities might be more important for learning newly introduced mathematics content (see also Geary, Hoard, Nugent, & Scofield, 2021; Geary, Hoard, Nugent, Ünal, & Scofield, 2020). Although these studies provide ample documentation of the relation between spatial and mathematical abilities, the relations between performance on specific spatial measures and overall mathematics achievement and in specific areas of mathematics are not fully understood.

To ensure a broad assessment of spatial abilities, we included measures that span a range of correlated but distinct skills (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006). The first measure is visuospatial working memory, which is predictive of gains in broad mathematics achievement during adolescence (Li & Geary, 2017) and might contribute to the development of spatial–numerical representations (Göbel, Calabria, Farne, & Rossetti, 2006). The second is a measure of visuospatial attention that is associated with generation of a mental number line and competence at estimating the relative magnitude of numerals (Geary, Scofield, Hoard, & Nugent, 2021; Longo & Lourenco, 2007; Zorzi et al., 2012); the measure was the Judgment of Line Angle and Position (JLAP) test (Collaer, Reimers, & Manning, 2007). The third measure is the extensively studied Mental Rotation Test (MRT; Peters et al., 1995), which involves the top-down manipulation of images, elicits broad engagement of visuospatial areas of the parietal cortex (Carpenter, Just, Keller, Eddy, & Thulborn, 1999), and is correlated with various mathematics outcomes (Halpern et al., 2007). For instance, performance on the MRT predicts accuracy in the solving of multistep word problems, overall performance on the SAT Math Test, and long-term success in math-intensive careers (Casey et al., 1997; Geary, Saults, Liu, & Hoard, 2000; Kell et al., 2013).

These relations are important because boys perform better than girls on all the visuospatial measures used in this study (Lauer, Yhang, & Lourenco, 2019; Voyer, Voyer, & Bryden, 1995; Voyer, Voyer, & Saint-Aubin, 2017). These patterns indicate that visuospatial skills are a plausible mechanism that could potentially compensate for boys' lower levels of engagement in classroom settings relative to those of girls. That said, we note that both attentional behavior in classroom settings and visuospatial abilities are malleable (DuPaul, Eckert, & Vilardo, 2012; Newcombe & Stieff, 2012; Tzuriel & Egozi, 2010; Uttal et al., 2013), and thus prior findings and those that might emerge in this study do not necessarily imply immutable sex differences.

The current study

The current study addressed the question of why boys perform as well as, or better than, girls in mathematics despite the sex difference favoring girls in in-class attentive behavior. The first goal was to determine whether teacher ratings—across grades and mathematics and English language arts (ELA) classrooms—predicted mathematics and, as a contrast, reading achievement in seventh and eighth grades. To evaluate the relative importance of in-class attentive behavior, we pitted it against the domain-general abilities of intelligence and working memory that are extensively studied and consistent predictors of mathematics achievement (Bull & Lee, 2014; Deary, Strand, Smith, & Fernandes, 2007; Geary, Nicholas, Li, & Sun, 2017; K. Lee & Bull, 2016). As shown below, in-class attentive behavior was a stronger predictor of mathematics achievement than intelligence and working memory, and girls were more engaged in classrooms than boys. Boys performed better than expected in mathematics given their level of classroom engagement, and as noted we explored whether this was related to their advantages in visuospatial skills.

Even though we had 2 years of data, individual differences in mathematics achievement (r = .85) and in-class attentive behavior (r = .74) were highly stable across grades and thus afforded little opportunity to assess developmental trends, likely due to the short timeframe. The one benefit was the ability to combine the same measures (e.g., digit span, in-class attentive behavior) across grades to create more reliable assessments of the associated constructs.

Method

Participants

The participants were 322 students (159 boys) enrolled in an ongoing longitudinal study conducted in collaboration with the Columbia Public Schools in Columbia, Missouri USA. They were recruited across two cohorts from a larger group of 1926 students who participated in an assessment of sixth-grade mathematical competencies (see Geary et al., 2019). All 1926 students were invited to join the longitudinal component of the study, and 342 of them and their parents did so. The 322 students included here completed all the seventh- and eighth-grade assessment sessions.

Of the 20 dropped students, 18 were administered the mathematics and reading measures (see below) in seventh grade (none of them completed all of the eighth-grade assessments), and there were no differences in achievement comparing these students with the 322 students used in these analyses (ps > .16). We decided to drop the seventh-grade data for these 18 students to keep the sample used in the analyses the same across grades and thus more comparable.

Demographic information was obtained through a parent survey. For the group of 322 students, 88% of them were non-Hispanic, 6% were Hispanic or Latino, and the ethnic status of the remaining students was unknown. The racial composition was 71% White, 14% Black, 3% Asian, 1% Native American, and 10% multiracial, with the remaining racial compositions unknown. As a comparison, in the school district from which the participants were recruited, 7% were Hispanic; 61% White, 20% Black, 5% Asian, 7% Hispanic, and 6% multiracial. For the current participants, parent-reported annual household income was distributed as follows: \$0 to \$24,999 (10%), \$25,000 to \$49,999 (18%), \$50,000 to \$74,999 (10%), \$75,000 to \$99,999 (21%), \$100,000 to \$149,999 (18%), and \$150,000+ (17%); the remaining parents did not respond to this question. Nearly two thirds (64%) of the students had at least one parent with a college degree. In addition, 16% of the families received food assistance, and 5% received housing assistance.

In-class attentive behavior

The Strength and Weaknesses of ADHD-Symptoms and Normal-Behavior (SWAN) measure (Swanson et al., 2012) was used to assess in-class attentive behavior. The items assess attentional deficits and hyperactivity, but the scores are normally distributed and based on the behavior of a typical student. The measure is more sensitive to variation in attentive behavior among the general school population than are commonly used diagnostic screeners for ADHD (Polderman et al., 2007). The inattentive component of such measures is more consistently related to academic outcomes than the hyperactivity component after the preschool years (Gray et al., 2017). Thus, the 9-item inattention subscale (e.g., "Gives close attention to detail and avoids careless mistakes") was distributed to the students' seventh- and eighth-grade mathematics and ELA teachers, who were asked to rate the behavior of the students relative to other students of the same age on a scale of 1 (*far below*) to 7 (*far above*).

Of the 322 students, 191 had ratings from two or three teachers, 128 had ratings from all four teachers, and 3 had a rating from only 1 teacher. Ratings were consistent across items (αs = .98) and mathematics and ELA teachers within and across grades (rs = .60–.75). For the students with all four ratings, a repeated-measures analysis of variance revealed an overall sex difference (see below), $F(1, 126) = 6.04$, $p = .015$, but no class (mathematics vs. ELA) by sex interaction, $F(1, 126) = 0.07$, $p = .797$, grade by sex interaction, $F(1, 126) = 1.08$, $p = .30$, or class by grade by sex interaction, $F(1, 126) = 0.01$, $p = .905$. Given the consistency of ratings across teachers and grades and the lack of interactions with student sex, we calculated one in-class attentive behavior score using the mean of available ratings across teachers and grades ($\alpha = .88$).

Standardized measures

Intelligence

Full Scale IQ was estimated using the Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), following procedures detailed in the manual.

Mathematics achievement

Mathematics achievement was assessed using the Numerical Operations subtest of the Wechsler Individual Achievement Test–Third Edition (Wechsler, 2009). For students of this age, the associated items included basic arithmetic and continued through fractions, algebra, geometry, and calculus, solved with pencil and paper.

Reading achievement

Two subtests of the Wechsler Individual Achievement Test–Third Edition (Wechsler, 2009) were used to assess reading achievement. The first was the Word Reading subtest, which assesses single word reading, beginning with one-syllable words and progressing to more complex vowel, consonant, and morphology types. The second was the Oral Reading Fluency subtest. Here, students read two passages (one at a time) under a time limit. Reading errors (e.g., added words, misstated words) were recorded by the experimenter, and the coding was later verified during a review of an audio-recording of the read passages. Completion time was independently verified using the same recording. The scores were reading accuracy [total word count − (total addition errors + total other errors)] and oral reading fluency {[(total word count–total other errors)/total completion time]*60}, which were highly correlated ($r = .68$, $p < .001$).

A reading fluency measure rather than comprehension measure was included because fluency measures are often used to assess reading difficulties (J. Lee & Yoon, 2017). This is important because one goal of the larger project (but not for this study) was to identify the sources of the comorbidity of mathematics difficulties and reading difficulties (Geary, Hoard, Nugent, Ünal, & Scofield, 2020).

Cognitive measures

The tasks included standard measures of short-term and working memory, which have been shown in prior studies to correlate with mathematics achievement, as well as various aspects of spatial ability. All the tasks were administered on iPads using customized programs developed through Inquisit by Millisecond (https://www.millisecond.com); manuals are available on the Open Science Framework (https://osf.io/qwfk6/). With the exception of the *N*-back task, which was administered only in seventh grade (due to time constraints), all tasks were administered in seventh and eighth grades.

We used the mean score across the two grades and estimated the reliabilities (ρ) of these summary scores using the Spearman–Brown prophecy formula applied to the test–retest correlations.

Digit span

Digits were presented at 1-s intervals, starting with three digits for the forward assessment and two digits for the backward assessment. Students' task was to recall the digit list in either a forward or backward manner, respectively, by tapping the digits on a display on the iPad screen. Students moved up to the next level if the response was correct. If the response was incorrect, the same level was presented a second time. If a consecutive error occurred, students moved down to a lower level. Each direction (forward and then backward) ended after 14 trials. The score was the highest digit span correctly recalled before making two consecutive errors at the same span length. Reliabilities for the current sample were adequate for both forward span (ρ = .67) and backward span (ρ = .73).

N-Back

Students completed an adaptive version of a single *N*-back task (Jaeggi et al., 2010). Students were first shown a "target" letter, followed by a sequence of 20 randomly determined stimulus letters (all consonants; 6 were a target letter and 14 were not). The task was to indicate whether the currently presented letter was a target by tapping a key or was not a target by not responding. The target letter could be the first stimulus presented ($N = 0$) or could be the same as the one that preceded it ($N = 1$) or the same as one presented in the two trials ($N = 2$) or three trials ($N = 3$) that preceded it.

For each trial, a letter was presented for 500 ms, followed by a 2500-ms blank screen. Students had 3000 ms to respond by tapping a key if the target was detected. After three 10-item practice blocks (levels $N = 0$ to $N = 2$), all participants started on level $N = 0$. Depending on performance, they moved up, stayed on the current level, or moved down a level for a total of five blocks (less than three errors: move up; three to five errors: repeat level; more than five errors: move down). Performance feedback (percentage correct) was displayed after each block. Hits (H), misses, false alarms (FA), and correct rejections were recorded and summarized by block. The score was (H − FA)/(total blocks). The estimated split-half reliability was .74 for the current sample.

Spatial span

The Corsi Block-Tapping Task (Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000) was used to assess forward spatial span. Students were presented with a display of nine squares that appeared to be randomly arranged. The squares "lighted up" in a predetermined sequence, and the task was to tap on the squares in the same order they were lit. The sequence length started at two squares and could increase to up to nine squares. Students had two attempts at each sequence length. If one of the sequences was recalled correctly, the next sequence level began; if both sequences at the same level were recalled incorrectly, the task was terminated. The score was the total number of correctly recalled sequences across the whole task (ρ = .69 for the current sample).

Spatial skills

Two tasks were used to assess spatial ability. The first was the JLAP test (Collaer et al., 2007), which assesses visuospatial attention (Benton, Varney, & Hamsher, 1978). Students were asked to match the angle of a single presented line to 1 of 15 line options in an array at the bottom center of the iPad screen. The 20 test items were presented one at a time, and students selected the matching angle by tapping it on the iPad touchscreen. Each stimulus was presented for up to 10 s, and when a selection was made a reaction time was recorded and the next stimulus was presented. The outcome was the number correct (ρ = .69 for the current sample).

The MRT (Peters et al., 1995) was the second spatial measure, which assesses the ability to manipulate images (Hegarty, 2018). On each trial, students viewed images of three-dimensional drawings of 10 connected cubes. For each trial, there was one target and four choice options, and the task was to select the two options that were rotations of the target figure. After four self-paced practice problems, students were presented with 24 problems in two blocks of 12 problems each (3 min per block). The score was the number of problems on which students chose both correct options (ρ = .83 for the current sample).

Procedure

The assessment schedule and the ages at each assessment are shown in Table 1. The 45-min assessments were done one on one in a quiet location in students' school; due to COVID-19-related school closings, spring assessments for eighth grade were completed online during a proctored video conference for 7 students. As shown in Table 1, the cognitive measures were administered during the first semester of each grade, and the remaining measures were administered during the second semester. During the spring of both grades, mathematics and ELA teachers completed the in-class attentive behavior survey.

Parents provided informed written consent, and assent was obtained from adolescents for all assessments. The University of Missouri institutional review board approved all methods included in this study.

Analyses

Hierarchical regression analyses were first used to determine whether in-class attentive behavior predicted achievement outcomes, with control of domain-general abilities (IQ, digit span forward, digit span backward, and *N*-back) and spatial abilities (Corsi, JLAP, and MRT). These analyses established the relative importance of in-class attentive behavior for predicting achievement outcomes; these outcomes and all predictors were centered ($M = 0$, $SD = 1$). We then examined the sex difference in in-class attentive behavior, focusing on (a) whether control of attentive behavior resulted in sex differences in mathematics and, as a control, reading achievement and (b) whether there was a sex by attentive behavior interaction. The latter was a validity check on the teachers' ratings. If the sex difference (see below) on these ratings is due to teacher bias favoring girls, then in-class attentive behavior should be a stronger predictor of girls' achievement outcomes than of boys' achievement outcomes, resulting in a significant interaction; the bias would result in more error in the assessment of boys' actual attentive behavior and thus a weaker relation between rated attentive behavior and achievement.

Finally, a mediation analysis using the lavaan package in R (Rosseel, 2012) was used to assess whether boys' advantages in Numerical Operations, after controlling for in-class attentive behavior, were mediated by their advantages on the spatial measures (see below). For these analyses, in-class attentive behavior and other significant predictors of achievement (e.g., intelligence) were used as control variables and the spatial measures that were significant in the initial regression analyses were assessed as potential mediators.

Table 1
Age of administration and timing of assessments.

Task name	Seventh grade		Eighth grade	
	Fall	Spring	Fall	Spring
Mean age at test[a]	153	157	164	168
In-class attentive behavior		x		x
Intelligence		x		
Digit span forward	x		x	
Digit span backward	x		x	
N-back	x			
Corsi	x		x	
Judgment of Line Angle and Position test	x		x	
Mental Rotation Test		x		x
Numerical Operations		x		
Word Reading				x
Oral Reading Fluency				x
Oral Reading Accuracy				

[a] Mean age is in months. Standard deviations range from 4.41 to 4.96 months.

Results

Mean scores and any associated sex differences are shown in Table 2, and correlations among the variables are shown in Fig. 1. Overall, boys and girls were more similar than different, although girls were more attentive in classroom settings ($d = -.34$) and boys had small to moderate advantages on the three spatial tasks (ds = .28–.56). The correlations show that all the spatial measures (i.e., Corsi, JLAP, and MRT) and in-class attentive behavior were more highly correlated with mathematics than with reading achievement, whereas digit span forward was more highly correlated with reading than with mathematics achievement.

In-class attentive behavior and academic outcomes

The results of the hierarchical regression models are shown in Table 3. The key finding was that in-class attentive behavior was a significant predictor of all achievement outcomes ($ps < .002$), controlling for domain-general and spatial abilities. Moreover, the magnitude of the effect for predicting Numerical Operations scores was more than double that found in the prediction of same-grade reading scores (e.g., $\beta = .37$, $\beta = .16$ for Numerical Operations and Word Reading in seventh grade). The other key finding was that JLAP and MRT scores were significant predictors of Numerical Operations scores in both grades, controlling for domain-general abilities and in-class attentive behavior. In contrast, these spatial measures were inconsistently related to reading outcomes; of the six effects, only two (MRT in predicting Word Reading and JLAP in predicting Oral Reading Accuracy) were significant.

Sex differences

The sex difference for in-class attentive behavior ($p = .002$, $d = -.34$) is shown in Fig. 2. As can be seen, the distribution of girls' scores was skewed toward the high end of the range, whereas boys' scores largely ranged from the middle to the high end. Despite boys' lower in-class attentive behavior and the importance of attentive behavior in the prediction of achievement scores, especially mathematics, there were no sex differences on any of the achievement tests ($ps > .50$).

Controlling for in-class attentive behavior resulted in sex differences favoring boys for the Numerical Operations subtest in seventh grade, $t(319) = 3.12$, $p = .002$, $d = .28$, and eighth grade, $t(319) = 3.20$, $p = .002$, $d = .27$, as well as for Word Reading, $t(319) = 2.14$, $p = .033$, $d = .22$. The sex differences in Oral Reading Accuracy and Oral Reading Fluency remained nonsignificant ($ps > .22$). The sex differences on the Numerical Operations subtest remained significant with control of domain-general abilities and in-class attentive behavior, as shown in Table 4; the spatial measures were not included in these regressions because we hypothesized that sex differences in mathematics outcomes would disappear

Table 2
Sex differences in mean scores.

Measure	Overall [*M* (*SD*)]	Boys [*M* (SD)]	Girls [*M* (*SD*)]	*t* Value	*p*	*d*[a]
In-class attentive behavior	4.89 (1.31)	4.66 (1.28)	5.11 (1.31)	−3.07	.002	−.34
Intelligence	104.96 (13.05)	105.43 (12.36)	104.50 (13.71)	0.64	.525	.07
Digit span forward	5.86 (0.99)	5.87 (1.02)	5.85 (0.95)	0.25	.802	.02
Digit span backward	4.72 (1.12)	4.71 (1.12)	4.73 (1.12)	−0.18	.858	−.02
N-Back	3.81 (0.76)	3.85 (0.80)	3.76 (0.71)	1.08	.280	.12
Corsi	8.59 (1.67)	8.82 (1.63)	8.35 (1.69)	2.55	.011	.28
Judgment of Line Angle and Position test	13.59 (2.86)	14.40 (2.75)	12.79 (2.76)	5.24	.000	.56
Mental Rotation Test	9.86 (4.32)	10.92 (4.65)	8.83 (3.70)	4.46	.000	.48
Numerical Operations, seventh grade	100.38 (18.67)	101.14 (18.59)	99.63 (18.78)	0.72	.470	.08
Word Reading, seventh grade	104.44 (13.27)	104.91 (13.12)	103.99 (13.44)	0.62	.536	.07
Numerical Operations, eighth grade	99.00 (18.19)	99.44 (18.61)	98.57 (17.82)	0.43	.669	.05
Oral Reading Fluency, eighth grade	103.49 (11.98)	103.14 (12.04)	103.83 (11.96)	−0.52	.603	−.06
Oral Reading Accuracy, eighth grade	92.68 (12.54)	92.56 (11.31)	92.80 (13.66)	−0.17	.865	−.02

[a] $d = M_m - M_f$/(pooled *SD*) where m = male and f = female.

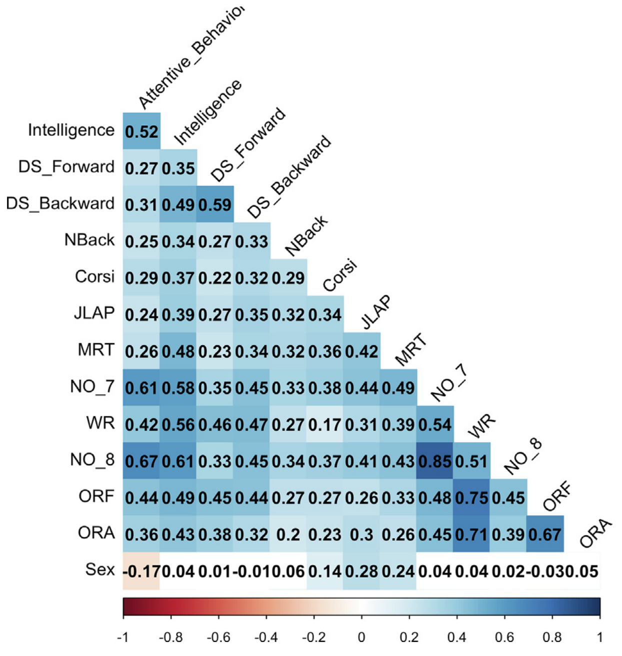

Fig. 1. Correlations among the variables: DS, digit span; JLAP, Judgment of Line Angle and Position; MRT, Mental Rotation Test; NO, Numerical Operations; WR, Word Reading; ORF, Oral Reading Fluency; ORA, Oral Reading Accuracy.

Table 3
Estimates from full regression models.

Predictor	Seventh grade				Eighth grade					
	Numerical Operations		Word Reading		Numerical Operations		Oral Reading Fluency		Oral Reading Accuracy	
	M (*SE*)	*p*	*M* (*SE*)	*p*	*M* (*SE*)	*p*	*M* (*SE*)	*p*	*M* (*SE*)	*p*
Intelligence	.11 (.05)	.039	.31 (.06)	.000	.19 (.05)	.000	.20 (.06)	.002	.23 (.06)	.000
Digit span forward	.03 (.05)	.600	.23 (.05)	.000	.02 (.05)	.658	.24 (.06)	.000	.20 (.06)	.000
Digit span backward	.11 (.05)	.048	.13 (.06)	.023	.09 (.05)	.076	.10 (.06)	.109	.14 (.06)	.026
N-Back	.04 (.04)	.410	.01 (.05)	.892	.06 (.04)	.188	.02 (.05)	.726	−.07 (.05)	.161
Corsi	.06 (.04)	.169	−.15 (.05)	.003	.04 (.04)	.368	.02 (.05)	.689	.02 (.05)	.675
JLAP test	.12 (.05)	.010	.02 (.05)	.648	.11 (.04)	.011	−.02 (.05)	.741	.16 (.05)	.003
Mental Rotation Test	.21 (.05)	.000	.15 (.05)	.003	.11 (.05)	.019	.08 (.06)	.125	−.05 (.05)	.358
Attentive behavior	.37 (.05)	.000	.16 (.05)	.002	.43 (.04)	.000	.20 (.05)	.000	.19 (.05)	.000

Note. JLAP, Judgment of Line Angle and Position.

with their control (see below). The regressions also show that the sex by attentive behavior interactions were nonsignificant for all academic outcomes, suggesting that teachers were not biased in their ratings.

The results so far indicate that with the control of in-class attentive behavior and domain-general abilities, boys had an advantage in mathematics but not in reading achievement. The mediation analyses tested the hypothesis that boys' advantage will be mediated by spatial abilities. For these analyses, we used JLAP and MRT scores as potential mediators because there are sex differences on these measures and they predict Numerical Operations scores in both grades, controlling for other factors. In the mediation analyses, the influence of intelligence, digit span backward, and in-class attentive behavior was controlled for seventh-grade Numerical Operations scores, and the influence of

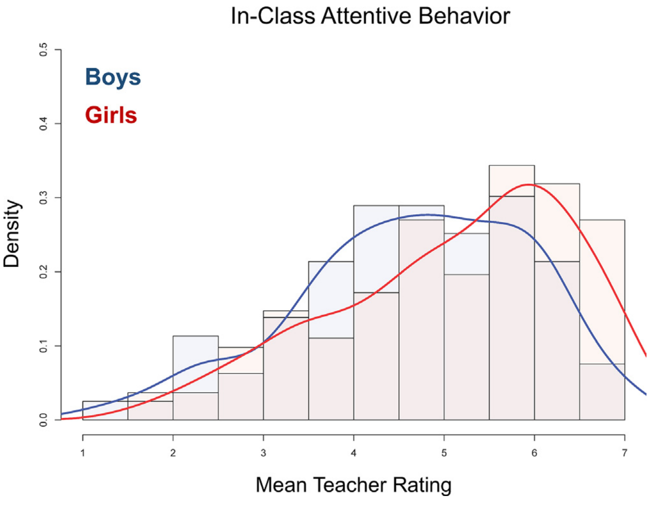

Fig. 2. Histograms and density plots of teacher ratings of boys' and girls' in-class attentive behavior.

Table 4
Estimates from full regression models without spatial measures.

Predictor	Seventh grade				Eighth grade					
	Numerical Operations		Word Reading		Numerical Operations		Oral Reading Fluency		Oral Reading Accuracy	
	M (*SE*)	*p*	*M* (*SE*)	*p*	*M* (*SE*)	*p*	*M* (*SE*)	*p*	*M* (*SE*)	*p*
Sex	−.19 (.08)	.024	−.09 (.09)	.296	−.18 (.08)	.024	.02 (.09)	.861	−.03 (.09)	.720
Intelligence	.22 (.05)	.000	.34 (.06)	.000	.25 (.05)	.000	.23 (.06)	.000	.24 (.06)	.000
Digit span forward	.03 (.05)	.602	.22 (.05)	.000	.02 (.05)	.650	.24 (.06)	.000	.21 (.06)	.000
Digit span backward	.16 (.06)	.005	.13 (.06)	.023	.13 (.05)	.013	.11 (.06)	.080	.16 (.06)	.010
N-Back	.09 (.05)	.038	.01 (.05)	.856	.09 (.04)	.030	.03 (.05)	.513	−.05 (.05)	.335
Attentive behavior	.37 (.07)	.000	.18 (.07)	.008	.47 (.06)	.000	.19 (.07)	.011	.13 (.07)	.068
Attentive Behavior * Sex	.06 (.08)	.495	−.06 (.09)	.500	−.02 (.08)	.799	.02 (.09)	.833	.13 (.09)	.158

Note. The sex contrast represents performance of girls relative to boys.

intelligence and in-class attentive behavior was controlled for eighth-grade scores; these covariates were included because they were significant predictors of these outcomes (Table 3).

As can be seen in Fig. 3 for seventh grade and in Fig. 4 for eighth grade, boys' advantage on the Numerical Operations subtest was fully mediated by the sex differences on the MRT and JLAP test. The results are the same, with only control of in-class attentive behavior.

We also ran a model that assessed an alternative explanation for the patterns in the data, specifically, whether the sex differences on the spatial tests mediated the sex difference in in-class attentive behavior. If true, then the sex difference in in-class attentive behavior is not fully separable from the sex differences on the spatial measures, which would confound any interpretation of the former. As can be seen in Fig. 5, all the paths and indirect effects are significant, but the sex difference in in-class attentive behavior is larger with control of the sex differences on the spatial measures, although the *c*' estimate of −0.76 is within the 95% confidence interval (CI) of the *c* estimate (CI = −1.03, −0.48). In any case, the patterns indicate that the sex differences in in-class attentive behavior and performance on the spatial measures are correlated but distinct phenomena.

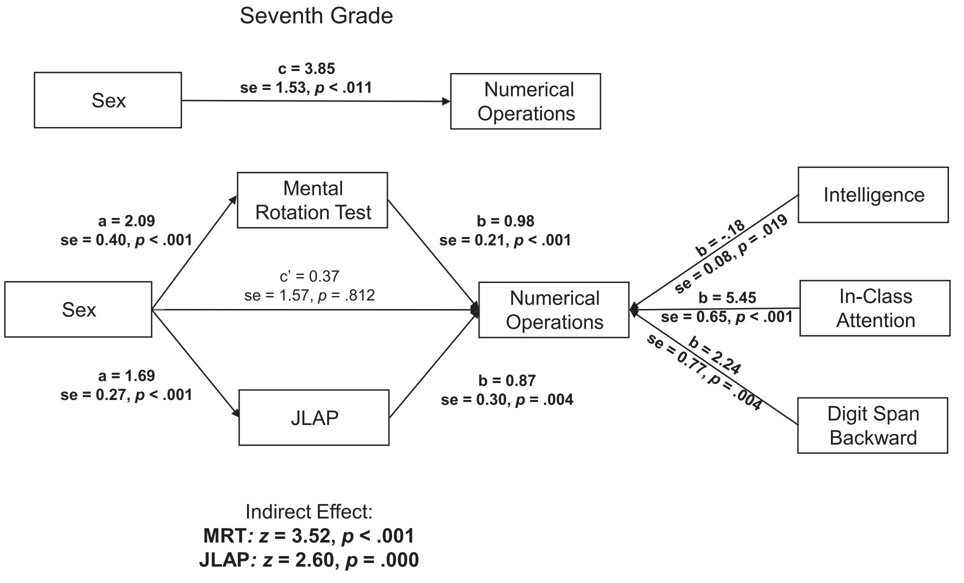

Fig. 3. Sex differences for Numerical Operations scores in seventh grade, controlling for intelligence, in-class attentive behavior, and digit span backward, were fully mediated by the sex differences on the Mental Rotation Test (MRT) and the Judgment of Line Angle and Position (JLAP) test. Significant effects are in bold. The results are the same (i.e., full mediation) with only control of in-class attentive behavior.

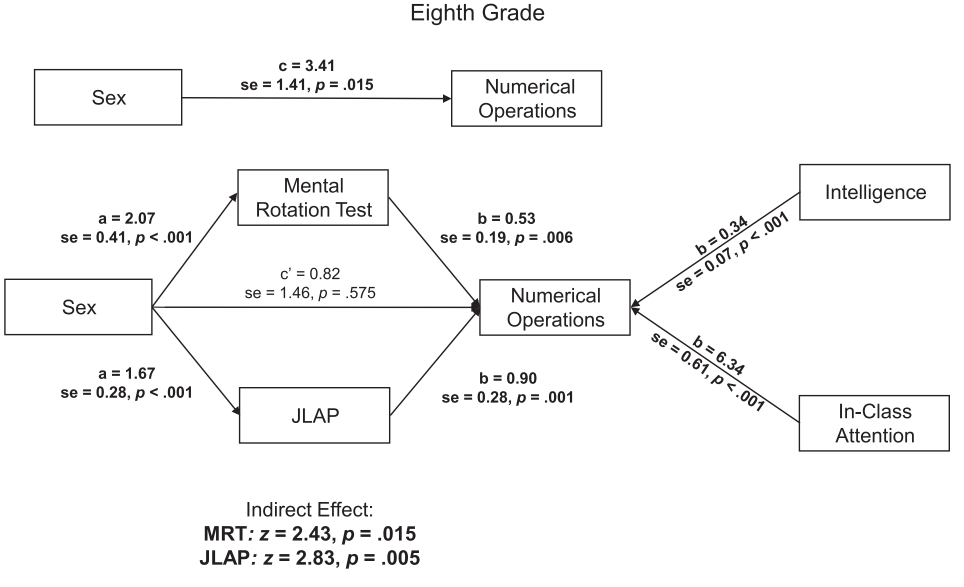

Fig. 4. Sex differences for Numerical Operations scores in eighth grade, controlling for intelligence and in-class attentive behavior, were fully mediated by the sex difference on the Mental Rotation Test (MRT) and the Judgment of Line Angle and Position (JLAP) test. Significant effects are in bold. The results are the same (i.e., full mediation) with only control of in-class attentive behavior.

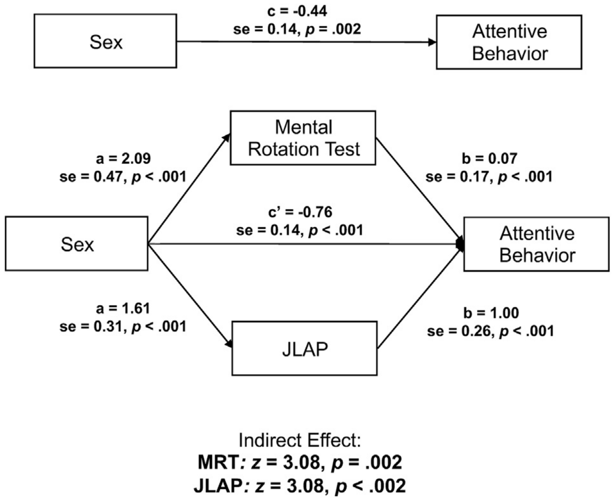

Fig. 5. Sex differences for in-class attentive behavior are not mediated by sex differences on the Mental Rotation Test (MRT) and the Judgment of Line Angle and Position (JLAP) test. Significant effects are in bold.

The final analyses assessed whether the strength of the relation between the MRT and JLAP test and mathematics achievement varied by sex. As shown in Fig. 6, the sex by MRT interaction was insignif-

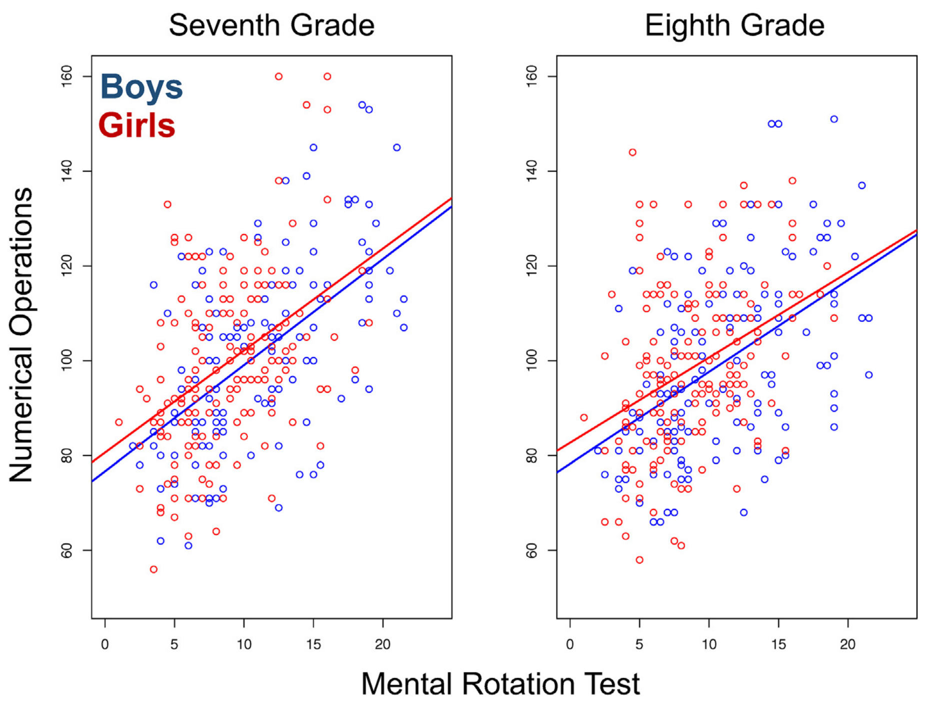

Fig. 6. Higher Mental Rotation Test scores predict higher overall Numerical Operations scores in both seventh and eighth grades and equally well for boys and girls.

icant in both seventh grade, $t(318) = 0.20$, $p = .845$, and eighth grade, $t(318) = 0.31$, $p = .756$. For both boys and girls, higher MRT scores were associated with higher Numerical Operations scores in seventh grade (*r*s = .56 and .42, respectively) and eighth grade (*r*s = .48 and .37, respectively). The same pattern emerged for the JLAP test; the sex by JLAP interaction was insignificant in both seventh grade, $t(318) = 0.17$, $p = .867$, and eighth grade, $t(318) = 1.21$, $p = .228$.

Discussion

To our knowledge, this study is the first to address the question of why the mathematics achievement of boys is just as good as, or better than, that of girls despite boys' relatively poor in-class attentive behavior. We organize the discussion around the sex difference in the latter and then turn to visuospatial skills and mathematics.

In-class attentive behavior

As found in many previous studies, ratings of in-class attentive behavior predicted mathematics and reading achievement (Duncan et al., 2007; Gray et al., 2017; Polderman et al., 2010; Stigler et al., 1987) and in fact were a more important predictor of mathematics achievement than the domain-general cognitive abilities of intelligence and working memory. Equally important, the ratings were consistent across teachers and the sex difference did not vary across grades or mathematics or ELA classrooms. The strength of the relations between in-class attentive behavior and achievement did not differ for boys and girls, indicating that the external validity of the ratings was the same across the sexes (see also Arnett et al., 2015). In other words, the ratings were not biased by the sex of the students and appeared to be capturing achievement-relevant behavior in classroom settings regardless of whether these were mathematics or ELA classrooms. Moreover, the magnitude of the sex difference in in-class attentive behavior for our sample ($d = -.34$) was highly consistent with prior studies of adolescents (*d*s = −.37 to −.32) (Lampert et al., 2004; Rietveld et al., 2004).

On the basis of these findings, we would have expected girls to outperform boys on all the achievement tests, but this was not the case. There were no significant sex differences in mathematics achievement, although boys' scores were slightly higher in both grades (*d*s = .08 and .05 in seventh and eighth grades, respectively), in keeping with a generally small sex difference on these types of tests (Hyde, Fennema, & Lamon, 1990; Lindberg, Hyde, Petersen, & Linn, 2010; Stoet & Geary, 2013). The lack of sex differences for the reading measures was surprising given that girls generally outperform boys in reading (Reilly, Neumann, & Andrews, 2019; Stoet & Geary, 2013). One potential reason is the complexity and foci of the assessments. Sex differences are typically assessed using composite reading measures or measures of reading comprehension, and the sex differences on these measures are typically larger (Reilly et al., 2019) than those found for word reading or oral reading fluency (Pargulski & Reynolds, 2017).

Whatever the reason, controlling for in-class attentive behavior resulted in an advantage for boys in mathematics achievement in both grades (*d*s = .27 and .28 in seventh and eighth grades, respectively) and an advantage in word reading ($d = .22$), but no effects on oral reading accuracy or fluency. The effect for word reading disappeared, however, with further control of domain-general abilities, but boys' advantage in mathematics remained. The finding for mathematics is particularly important given the primacy of in-class attentive behavior for the prediction of mathematics achievement for both boys and girls. The pattern suggests that boys' lower overall engagement in classroom settings is modestly compromising their mathematics learning and might have effects on their reading competencies, but any such effects are less consistent than those found for mathematics (Stigler et al., 1987).

Visuospatial skills and mathematics

Boys' advantages on the visuospatial measures are also consistent with previous findings, although the magnitudes of the effect sizes are smaller than those often found during adulthood (Collaer &

Nelson, 2002; Voyer et al., 1995, 2017). The latter is likely because the magnitude of the sex differences on visuospatial measures increases from childhood to early adulthood (Geary, 2021; Lauer et al., 2019) and thus is often smaller during early adolescence than during adulthood.

The key finding here is not the sex differences but rather the mediation of the sex differences in mathematics achievement after controlling for in-class attentive behavior. Despite the importance of in-class attentive behavior in the prediction of mathematics achievement (Stigler et al., 1987), boys were performing better in mathematics than might be expected based on their attentiveness in classroom settings. In other words, if there are small or no sex differences at the mean for mathematical competencies, then girls should have had higher mathematics achievement scores given their higher attentiveness in mathematics classrooms. The better-than-expected performance and the sex differences that emerged from this were explained by the complex spatial abilities captured by the MRT (Peters et al., 1995) and the JLAP test (Collaer et al., 2007). The findings for the MRT are consistent with prior studies (Casey et al., 1997; Geary et al., 2000), and the findings for the JLAP test are consistent with studies showing that visuospatial attention contributes to the processing of numerical magnitudes (Geary et al., 2020; Longo & Lourenco, 2007).

However, the specific reasons for the relations between MRT and JLAP performance and mathematics achievement cannot be determined from these results. Nevertheless, as shown in Fig. 6, overall mathematics achievement was linearly related to MRT and JLAP scores for both boys and girls. There are two important takeaways from these relations. The first is that better spatial abilities predict better performance on more difficult items on the achievement test. Students are generally less familiar with these items than the less difficult ones, and therefore this result is consistent with Mix et al.'s suggestion (Mix, 2019; Mix et al., 2016), specifically, that spatial abilities may be most critical for learning newly introduced mathematics content. The combination of studies suggests that visuospatial abilities may facilitate the representation or understanding of newly introduced mathematical content, but how this might occur cannot be determined from the current study.

The second takeaway is that the relation between MRT and JLAP performance and mathematics achievement is the same for boys and girls; well-developed visuospatial skills are important for both sexes. Boys' better-than-expected mathematics achievement is not because the relation between visuospatial abilities and mathematics differs for boys and girls but rather because there are more boys than girls with well-developed visuospatial abilities. The same is true for in-class attentive behavior. Engagement in classroom settings is associated with the mathematics achievement of both sexes. Girls seem to benefit more from this engagement, but this is only because there are more girls than boys at the high end of in-class attentive behavior.

Limitations and conclusion

The correlational nature of the study precludes causal statements. Although we assessed a broader array of potential predictors of achievement than is typical in this type of study, there may be other factors that we did not include that could change the pattern of results. Moreover, we were unable to control for classroom-level effects that may have influenced the pattern of findings. Despite these limitations, the study provides a more reliable assessment of the relation between visuospatial performance and mathematics achievement and associated sex differences than is typical in this literature with the multiple assessments of working memory, spatial abilities, in-class attentive behavior, and achievement. The framing of the study in terms of sex differences in in-class attentive behavior as related to achievement is unique (to our knowledge) and adds nuance to our understanding of these relations.

Acknowledgments

The study was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD087231) and from the National Science Foundation (DRL-1659133). We thank Dana Hibbard of Columbia Public Schools, as well as the staff, faculty, and students and their families for their assistance with the logistics of the study. We thank Kristin Balentine, Mandar Bhoyar, Amanda Campbell, Maria Ceriotti, Felicia Chu, Anastasia Compton, Alexis Currie, Kaitlynn Dutzy, Amanda Evans, James Farley, Amy Jordan, Joseph LaMendola, Bradley Lance, Kate Leach, Joshua McEwen, Kelly Mebruer, Heather Miller, Sarah Peterson, Nicole Reimer, Laura Roider, Brandon Ryffe, Logan Schmidt, and Melissa Willoughby for their assistance with data collection and processing.

References

Arnett, A. B., Pennington, B. F., Willcutt, E. G., DeFries, J. C., & Olson, R. K. (2015). Sex differences in ADHD symptom severity. *Journal of Child Psychology and Psychiatry, 56*, 632–639.

Benton, A. L., Varney, N. R., & Hamsher, K. D. (1978). Visuospatial judgment: A clinical test. *Archives of Neurology, 35*, 364–367.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*, 36–41.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W., & Thulborn, K. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience, 11*, 9–24.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*, 669–680.

Ceci, S. J., & Williams, W. M. (2010). *The mathematics of sex: How biology and society conspire to limit talented women and girls*. New York: Oxford University Press.

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*, 2–11.

Cheung, C. N., Sung, J. Y., & Lourenco, S. F. (2019). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research Psychologische Forschung, 84*, 2000–2017.

Collaer, M. L., & Nelson, J. D. (2002). Large visuospatial sex difference in line judgment: Possible role of attentional factors. *Brain and Cognition, 49*, 1–12.

Collaer, M., Reimers, L., & Manning, S. (2007). Visuospatial performance on an internet line judgment task and potential hormonal markers: Sex, sexual orientation, and 2D:4D. *Archives of Sexual Behavior, 36*, 177–192.

Daucourt, M. C., Erbeli, F., Little, C. W., Haughbrook, R., & Hart, S. A. (2020). A meta-analytical review of the genetic and environmental correlations between reading and attention-deficit/hyperactivity disorder symptoms and reading and math. *Scientific Studies of Reading, 24*, 23–56.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Sexton, H. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446.

DuPaul, G. J., Eckert, T. L., & Vilardo, B. (2012). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis 1996–2010. *School Psychology Review, 41*, 387–412.

Friso-van den Bos, I., & van de Weijer-Bergsma, E. (2020). Classroom versus individual working memory assessment: Predicting academic achievement and the role of attention and response inhibition. *Memory, 28*, 70–82.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29–46.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., & Hamlett, C. L. (2014). Sources of individual differences in emerging competence with numeration understanding versus multidigit calculation skill. *Journal of Educational Psychology, 106*, 482–498.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*, 229–284.

Geary, D. C. (2021). *Male, female: The evolution of human sex differences* (3rd ed.). Washington, DC: American Psychological Association.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, H. D. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE, 8*(1), e54651.

Geary, D. C., Hoard, M. K., Nugent, L., Chu, F. W., Scofield, J. E., & Hibbard, D. F. (2019). Sex differences in mathematics anxiety and attitudes: Concurrent and longitudinal relations to mathematical competence. *Journal of Educational Psychology, 111*, 1447–1461.

Geary, D. C., Hoard, M. K., Nugent, L., & Scofield, J. E. (2021). In-class attention, spatial ability, and mathematics anxiety predict across-grade gains in adolescents' mathematics achievement. *Journal of Educational Psychology, 113*, 754–769.

Geary, D. C., Hoard, M. K., Nugent, L., Ünal, Z. E., & Scofield, J. E. (2020). Comorbid learning difficulties in reading and mathematics: The role of intelligence and in-class attentive behavior. *Frontiers in Psychology, 11*, 572099. https://doi.org/10.3389/fpsyg.2020.572099.

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*, 680–693.

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*, 337–353.

Geary, D. C., Scofield, J. E., Hoard, M. K., & Nugent, L. (2021). Boys' advantage on the fractions number line is mediated by visuospatial attention: Evidence for a parietal–spatial contribution to number line learning. *Developmental Science, 24*, e13063.

Göbel, S. M., Calabria, M., Farne, A., & Rossetti, Y. (2006). Parietal rTMS distorts the mental number line: Simulating "spatial" neglect in healthy subjects. *Neuropsychologia, 44*, 860–868.

Gray, S. A., Dueck, K., Rogers, M., & Tannock, R. (2017). Qualitative review synthesis: The relationship between inattention and academic achievement. *Educational Research, 59*, 17–35.

Halpern, D., Benbow, C., Geary, D. C., Gur, R., Hyde, J., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*(1), 1–51.

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review, 27*, 465–482.

Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review, 25*, 1212–1219.
Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*, 151–176.
Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.
Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between *n*-back performance and matrix reasoning—Implications for training and transfer. *Intelligence, 38*, 625–635.
Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24*, 1831–1836.
Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology, 7*, 252–258.
Lampert, T. L., Polanczyk, G., Tramontina, S., Mardini, V., & Rohde, L. A. (2004). Diagnostic performance of the CBCL–Attention Problem Scale as a screening measure in a sample of Brazilian children with ADHD. *Journal of Attention Disorders, 8*, 63–71.
Lauer, J. E., & Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychological Science, 27*, 1291–1298.
Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: A meta-analytic review. *Psychological Bulletin, 145*, 537–565.
Lee, J., & Yoon, S. Y. (2017). The effects of repeated reading on reading fluency for students with reading disabilities: A meta-analysis. *Journal of Learning Disabilities, 50*, 213–224.
Lee, K., & Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology, 108*, 869–882.
Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PLoS ONE, 12*(2), e172046.
Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135.
Longo, M. R., & Lourenco, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia, 45*, 1400–1407.
Mix, K. S. (2019). Why are spatial skill and mathematics related?. *Child Development Perspectives, 13*, 121–126.
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*, 1206–1227.
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*, 465–492.
Murray, A. L., Booth, T., Eisner, M., Auyeung, B., Murray, G., & Ribeaud, D. (2019). Sex differences in ADHD trajectories across childhood and adolescence. *Developmental Science, 22*, e12721.
Newcombe, N. S., & Stieff, M. (2012). Six myths about spatial thinking. *International Journal of Science Education, 34*, 955–971.
Pargulski, J. R., & Reynolds, M. R. (2017). Sex differences in achievement: Distributions matter. *Personality and Individual Differences, 104*, 272–278.
Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39–58.
Polderman, T. J., Boomsma, D. I., Bartels, M., Verhulst, F. C., & Huizink, A. C. (2010). A systematic review of prospective studies on attention problems and academic achievement. *Acta Psychiatrica Scandinavica, 122*, 271–284.
Polderman, T. J., Derks, E. M., Hudziak, J. J., Verhulst, F. C., Posthuma, D., & Boomsma, D. I. (2007). Across the continuum of attention skills: A twin study of the SWAN ADHD rating scale. *Journal of Child Psychology and Psychiatry, 48*, 1080–1087.
Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., & Todd, R. D. (2010). Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 217–228.
Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist, 74*, 445–458.
Rietveld, M. J., Hudziak, J. J., Bartels, M., Van Beijsterveldt, C. E. M., & Boomsma, D. I. (2004). Heritability of attention problems in children: Longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry, 45*, 577–588.
Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*, 1301–1308.
Rivera-Batiz, F. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources, 27*, 313–328.
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.
Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1987). Mathematics classrooms in Japan, Taiwan, and the United States. *Child Development, 58*, 1272–1285.
Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE, 8*(3), e57988.
Stoet, G., & Geary, D. C. (2020). Gender differences in the pathways to higher education. *Proceedings of the National Academy of Sciences of the United States of America, 117*, 14073–14076.
Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A., ... Wigal, T. (2012). Categorical and dimensional definitions and evaluations of symptoms of ADHD: History of the SNAP and the SWAN rating scales. *International Journal of Educational and Psychological Assessment, 10*, 51–70.
Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81*, 1417–1430.
Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*, 352–402.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). I. Spatial skills, their development, and their links to mathematics. *Monographs of the Society for Research in Child Development, 82*(1), 7–30.
Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270.
Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual–spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24*, 307–334.
Wai, J., Hodges, J., & Makel, M. C. (2018). Sex differences in ability tilt in the right tail of cognitive abilities: A 35-year examination. *Intelligence, 67*, 76–83.
Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: PsychCorp, Harcourt Assessment.
Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Psychological Corporation.
Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: A meta-analytic review. *Neurotherapeutics, 9*, 490–499.
Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities, 33*, 179–191.
Zorzi, M., Bonato, M., Treccani, B., Scalambrin, G., Marenzi, R., & Priftis, K. (2012). Neglect impairs explicit processing of the mental number line. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00125.