Received: 10 May 2021 | Revised: 22 September 2021 | Accepted: 29 October 2021

DOI: 10.1111/ssm.12506

WILEY SSMA

RESEARCH PAPER – MATHEMATICS EDUCATION

Racial and gender disparities in elementary mathematics

Phillip A. Boda[1] | **Katherine James**[2] | **Jose Sotelo**[2] | **Steven McGee**[3] | **David Uttal**[2]

[1]University of Illinois at Chicago, Chicago, Illinois, USA

[2]Northwestern University, Evanston, Illinois, USA

[3]The Learning Partnership, Chicago, Illinois, USA

Correspondence
Phillip A. Boda, University of Illinois at Chicago, Chicago, IL, USA.
Email: paboda@uic.edu

Funding information
Northwestern University, Grant/Award Number: 1759360

Abstract

Disparities in Mathematics performance have been shown across race and gender for decades, although little research has reported the race by gender nexus in terms of disparity. In turn, research done to ameliorate these disparities have assumed that these differences are primarily among student populations who need remedial learning improvement. In this research, we test these assumptions by using multinomial regression analyses and adjusted mean comparisons with bias-corrected effect sizes from a data set of ~20,000 2nd grade students in Chicago. Further, we contribute to previous elementary Mathematics research by illuminating the powerful impacts that spatial reasoning integrations into Mathematics curriculum and pedagogy can have for all students' learning. Our findings provide empirical support that even after accounting for the effects of school composition, racial disparities exist in the beginning and persist throughout the year, among 2nd grade students. However, upon disaggregating students by race, gender, and median split, such disparities are greatest among students that are above the median. Our results encourage quantitative research analyses that explore ways to reduce Mathematics disparities to sharpen their approaches to be more sensitive to the achievement levels of students among such equity inquiries.

KEYWORDS

disparity, elementary math performance, multinomial regression, spatial reasoning

1 | INTRODUCTION

Developing novel learning opportunities and instruction in Mathematics at the elementary level is imperative to meet the needs of all students (Nicol & Crespo, 2005; Ottmar et al., 2013; Pianta et al., 2008). Longitudinal analyses across national, local, and comparative data sets highlight that growth and disparity in Mathematics begins and persists before secondary school (cf. Cameron et al., 2015; Kuzmina & Ivanova, 2018; Lee, 2010; Scammacca et al., 2020). This suggests that elementary education is one context to examine when differences in Mathematics achievement might emerge and point to approaches to ameliorate these disparities early. Short interventions have demonstrated benefits that last for months (Brisson et al., 2017), while cross-disciplinary impacts have shown powerful promise for students who often have the largest differences in Mathematics achievement (Shin et al., 2013). These opportunities are mediated, though, by how teachers envision what support they need professionally, their perceptions of student capabilities across race and gender, and what constitutes equitable Mathematics education goals (cf. Brand et al., 2006; Ford, 1994; Ottmar et al., 2013; Ross, 1995; Sparks & Pole, 2019).

The research presented here corroborates and contributes to past analyses that illuminate how students from diverse backgrounds are supported, or not, to achieve in Mathematics at the K-5 grade levels (cf. Brand et al., 2006;

Ottmar et al., 2013; Reardon & Galindo, 2009). We also investigate whether under-represented youth in science, technology, engineering, and mathematics (STEM) disciplines more broadly perform worse among Mathematics topics (Bachman et al., 2015). Through this research, we expand on analyses done that have showcased the impact of Mathematics tracking in relation to student opportunities (Ngo & Velasquez, 2020), sharpening this approach to focus more on what kind of disparities emerge and persist in order to ameliorate differences in Mathematics performance among all students. Inevitably, through our analyses we provide a highly sensitive approach to reveal the specific level of achievement where such disparities emerge and persist in relation to urban elementary Mathematics student performance in the third largest urban school district in the United States: Chicago, IL.

To this end, as the Common Core State Standards (CCSS) are used in Illinois, and these CCSS are aligned with assessments used by Chicago Public schools to measure Math Learning (NWEA, 2019; Set, 2018), we analyze these standards and their coinciding Mathematics Performance scales around topics in elementary Mathematics that are uniquely framed to benefit from the integration of spatial reasoning. Spatial reasoning was used as a specific theoretical frame for our inquiries given that studies have consistently confirmed its relationship with Mathematics achievement (cf. Geer et al., 2019; Gilligan et al., 2019; Hawes & Ansari, 2020; Young et al., 2018). There is also evidence to suggest that improving spatial reasoning may help ameliorate disparities across race and gender (Hadi-Tabassum, 2017; Lauer et al., 2019). Finally, integration of spatial reasoning has been shown to support equitable Mathematics learning across various achievement levels, and specifically for those populations from our data set where these disparities are most prevalent (i.e., high achievers; Rutherford et al., 2014). In sum, this work (1) highlights where spatial reasoning could be integrated in the Common Core State Standards for Mathematics; and (2) analyzes district data among all students' Mathematics performance across one school year at three points to report if disparities in Mathematics achievement exist and/or persist along spatially aligned content areas.

2 | BACKGROUND LITERATURE

Spatial reasoning plays a vital role in how we interact with our physical and mental world and is defined as “the ability to generate, retain, retrieve, and transform well-structured visual images” (Lohman, 1993, p.97). In other words, spatial reasoning focuses on the locations of objects in space, their shapes and movement, and relationships among these factors (Uttal & Cohen, 2012; Sauter et al., 2012). Most importantly, there is a wealth of evidence that spatial reasoning skills are critical to STEM learning (Stieff & Uttal, 2015), and that it plays an important role in early Mathematics skills, such as one to-one correspondence, number line knowledge, and geometry (Clements et al., 1999). Spatial skills are also associated with greater academic, occupational, and creative success in STEM domains (Kell et al., 2013; Wai et al., 2009).

However, important disparities exist in spatial reasoning skills. One of the most consistent findings is that males have better spatial reasoning than females (Linn & Petersen, 1985; Maeda & Yoon, 2013). This gender disparity has been observed throughout the lifespan and has been detected as early as infancy (cf. Moore & Johnson, 2008; Quinn & Liben, 2008). Although little work has been done to explore racial and socio-economic differences in spatial reasoning, there is evidence of achievement gaps in Mathematics among Black and White students (e.g., Lee & Wong, 2004), as well as socio-economic disparities impacting early Mathematics skills (e.g., Goldin & Katz, 2009). These disparities can invariably lead to larger gaps in education and career achievement in the future (Bradley & Corwyn, 2002). Improving spatial reasoning skills, thus, offers benefits for later STEM opportunities (Uttal et al., 2013; Halpern et al., 2007).

Below we unpack analyses of such disparities a bit further, along with examples from the literature that suggest that these differences in spatial reasoning can be ameliorated by design.

2.1 | Developing spatial reasoning: Prominence and impact of gender research studies

One of the most highly cited meta-analyses around differences in gender performance on spatial reasoning tasks (i.e., mental rotation) illuminates the emergence and persistence of disparities linked explicitly to its emergence in elementary school (Lauer et al., 2019). In this work, the authors tease out the nuances of these gender disparities around spatial reasoning among 128 studies estimating 303 effect sizes. Their conclusions paint a clear picture about the emergence and prominent increase of spatial reasoning between genders across schooling years:

> Spatial skills during childhood are predictive of later academic, professional, and creative success in STEM fields … [and] male advantage in mental rotation contributes to gender disparities in STEM attainment by early adolescence … interventions that specifically target factors implicated in girls'

> underperformance on mental rotation tasks (e.g., spatial anxiety, strategy use) could be effective in reducing gender disparities during development. Because these factors have been linked to individual differences in mental rotation performance among girls and boys alike, such interventions could not only ameliorate gender differences in children's mental rotation skills but also foster spatial development more broadly. (Lauer et al., 2019, p. 551–552)

Corroborating this meta-analytic finding, Levine et al. (2016) previously confirmed these claims and further teased out that the predictive success in STEM disciplines was similar among variable levels of achievement. These findings led us to explicitly think about within-group variance related to gender differences across races and the stratification of estimates based on performance among elementary students in Chicago Public Schools. More importantly, we also sought to leverage previous impact studies that sought to ameliorate these gender differences in terms of how we could make suggestions to integrate spatial reasoning in curriculum standards.

For example, a study by Lowrie et al. (2017) suggest that when designed collaboratively with elementary teachers, spatial reasoning infusions into existing Mathematics curriculum significantly and substantially impact both young students' spatial reasoning performance and their Mathematics learning. Additionally, a key take-away from this work is that there were no differential gains between genders on either measure of spatial reasoning or Mathematics learning, with these estimated differences well above significance thresholds ($p > .7$). This curricular co-design involved both physical artifacts and computer-supported manipulatives and did not increase the time spent on Mathematics instruction, which suggests that appropriate and powerful impacts can occur to ameliorate spatial reasoning and Mathematics learning disparities concurrently, while also not creating greater demand on classroom teachers in terms of curriculum scope and sequence. Further aligning with this single study, another meta-analysis of impact studies among young children designed to improve spatial reasoning through physical artifacts, visual prompting, and use of gesture supports that the effects of such impact studies are greater for girls ($g = 0.909$) than for boys ($g = 0.686$), which illustrates the power of designed support for such skills and thinking in relation to ameliorating disparities (Yang et al., 2020). From these works, we feel confident to claim that spatial reasoning is important for elementary Mathematics instruction, but also realize the lacking prevalence of such tasks among standards.

Research and policy efforts to develop and test explicit spatial reasoning instruction have been limited and, despite its importance, "[skill] in spatial thinking is presumed throughout the K-12 curriculum but is formally and systematically taught nowhere." (National Research Council, 2006, p. 131). Importantly, though, there is evidence that spatial reasoning can be improved with consistent practice and curriculum integration (Uttal, 2000; Newcombe, 2010), as we reported above. In turn, much of the work that has been done has focused on training specific skills, like mental rotation, in a laboratory setting (Sorby et al., 2013), which suggests that more work is needed to understand the effects of spatial reasoning instruction and bridge the gap between these lab studies and classroom-based learning and instruction (Hawes et al., 2015). To further elaborate on one example, an approach that has shown great promise for improving young children's skills is called *Spatial Thinking in Context* (StC; Newcombe, 2013).

This StC approach emphasizes spatial thinking as a series of practices for learning and teaching in STEM fields, with particular relevance to Mathematics instruction. For example, graphing and visualization techniques can be taught in the context of a variety of Mathematics topics, and these kinds of visualizations can help students begin to learn about graphing and graphical representations being part of Mathematical practices as recommended by the Common Core Math Standards. In one example, using area to indicate number and amount, either through manipulatives or representations, can help improve spatial reasoning by giving a concrete, spatial grounding to abstract Mathematics concepts (Newcombe, 2010). In geometry, students can also use blocks and other manipulatives to create models of key shapes and patterns, such as symmetry.

The challenge of this current work is to analyze existing opportunities to integrate such spatialized activities in the current widespread educational standards, the Common Core State Standards, and to understand the extent to which gender and racial disparities exist in young learners' Mathematics performance related to topics that are unique areas of Mathematics where spatial reasoning plays a crucial role. Given the known relationship between spatial ability and Mathematics performance (Holmes et al., 2008; Rasmussen & Bisanz, 2005), the goal of this work is to use Mathematics assessment data as a lens into spatial reasoning skills and integrate explicit spatial reasoning instruction to reduce early Mathematics disparities. We focus on Mathematics because a spatialized approach to early Mathematics instruction can have a transformative effect on learning. The research presented here, aligned with these goals, is guided by the following research questions:

1. What K-2 Mathematics standards contain opportunities to integrate spatial reasoning instruction?
2. Among these Mathematics topics, do gender and racial disparities exist? And do they persist?

3 | METHODS

This work is part of a researcher-practitioner partnership (RPP) between Chicago Public Schools (CPS), the Spatial Intelligence and Learning Center at Northwestern University, and The Learning Partnership focused on how explicit instruction in spatial reasoning in the primary grade band can contribute to reductions in variation in STEM outcomes for low-income, minority students in CPS. At the heart of the problem of practice for the RPP is research evidence that a significant proportion of the variance in STEM outcomes can be explained by spatial reasoning abilities, yet spatial reasoning is not routinely taught in K-12 settings. With seed funding from the Northwestern CPS Rapid Impact Grants Program, the spatial reasoning RPP conducted a needs assessment that clarified the problem of practice and identified opportunities for infusing spatial reasoning into the most prevalent math, science, and computer science curricula in CPS. Given the lack of high stakes assessments in the K-2 grade band, the CPS partners reported that the K-2 grade band received less attention than the other grade bands. Thus, this RPP focused on K-2 to help to fill that gap for CPS. Conversations amongst the partners led to a focus on three strands of work for the needs assessment: (1) an analysis of the variation in assessment performance at the initial grade level assessed in math and science (2nd grade); (2) a survey of K-2 CPS teachers to identify opportunities for engaging teachers in a co-development process; and (3) an analysis of the opportunities for infusing spatial reasoning into the most prevalent math, science, and computer science curricula in CPS. The goal of the needs assessment was to inform the partners on what concepts to focus on, where in the curricula to focus, and how to engage K-2 teachers in a co-design process.

3.1 | Data measures

Chicago Public Schools follow the Mathematics Common Core State Standards. We analyzed the four thematic Mathematics standards for each grade in the K-2 band to determine opportunities to integrate spatial reasoning: (a) Algebra, (b) Base 10, (c) Measurement & Data, and (d) Geometry. We then used the Northwest Evaluation Association's (NWEA) open-source data of the Measures of Academic Progress (MAP) Math Assessment for 2nd graders to explore any racial or gender disparities, and any disparities at their identity nexus (Race by Gender). The MAP Math Assessment aggregates scores across four Mathematics topics also reflected in the Common Core: (a) Operations and Algebraic Thinking; (b) Numbers and Operations; (c) Measurement and Data; (d) Geometry. Given the overlap in topic areas, analyzing both the MAP assessment data and the Common Core State Standards enabled us to examine correspondences between our two goals and areas of spatial integration. Moreover, NWEA has reported that their Math Assessment is aligned to assess these CCSS for Illinois students (NWEA, 2019; Set, 2018).

The MAP assessment was administered to all Chicago public school students during the 2018–2019 school year at three points: (a) Fall, (b) Winter, and (c) Spring. We analyzed each time point for differences among race and gender. Meaning, we compared students' racial categories in terms of their group performance on the MAP math assessment for 2nd graders analyzing any racial differences between Asian, Hispanic, Other, and Black students to white male students, and then explored any differences among genders within each racial category to give a profile analysis at each time point. This was done first at the aggregate score across all four Mathematics topics, then sharpened to examine if disparities exist and persist in two of the sub-score Mathematics topics that we found to be most appropriately aligned to having spatial reasoning integrated into the CCSS standards we analyzed. We unpack and tease out the nuances to this analytic approach and its limitation below for clarity of making white males the referent.

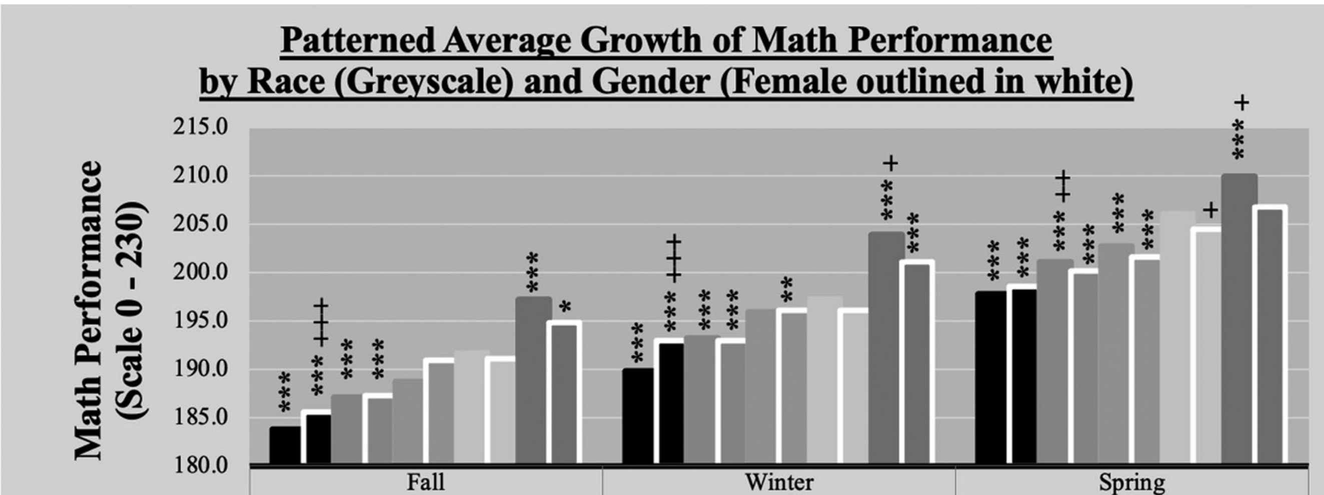
3.2 | Data analysis

To answer Research Question 1, the authors worked collaboratively to analyze the standards provided by the Common Core State Standards in Mathematics. Through an iterative analysis (elaborated more in the Findings section), we determined the extent to which these standards could draw from the integration of spatial reasoning based on the above literature. To answer Research Question 2, we conducted multinomial logistic regression analyses with and without school factors using white students as the referent category to compare if any differences existed between racial performance. We then calculated the Estimated Marginal Means of each sub-group of students (race interactions with gender) at all three time points. These values were also adjusted for the percentage of Free and Reduced lunch student qualifiers and English language learners at the school-level. These group means report adjusted Mathematics performance values while accounting for differences in sample sizes (Bonferroni Correction) to illuminate if any significant differences exist within racial groups for gender and compared to white males.

We also provide effect sizes (Hedge's *G*) for significant differences among races and genders that persist throughout the year to provide estimates of the magnitude of disparities to further bolster our claims. These effect sizes gave a bias-corrected estimate accounting for differences in variance that could affect our claims. These specific comparative analyses using white males as the referent category were done to examine how students perform in relation to what scholars denote as the "normative center of schools" (Leonardo & Broderick, 2011), which includes curriculum standards and performance. This analytic technique is further supported by the prominence in STEM education having much of the curriculum, pedagogies, and assessments historically and currently used that continue to support white male students to achieve and excel with little attention to marginalized student populations and their academic needs (cf. Garcia-Olp et al., 2019; Miles et al., 2019; Sengupta-Irving & Vossoughi, 2019; Visintainer, 2020). Indeed, while some may envision we are succumbing to "gap-gazing," we trouble this claim below.

3.2.1 | Novelty of our analytic approach, and its limitations

Gutiérrez's (2008) encouraged the field of Mathematics Education to "decenter" math achievement comparisons across racial disparities to focus instead on "excellence" and "gains" *within* historically marginalized groups. We want to emphasize that our current analysis was first engaged with an analysis that explored "growth" in relation to students' unmet socio-cultural support needs around elementary Mathematics teaching and learning. From that analysis, we found that all students do grow significantly in their Mathematics learning across the NWEA assessment (Figure 1 and Table 2 showcase this growth across all race and gender scores). Additionally, we constructed trend lines of aggregate NWEA Mathematics growth across our racial groups compared to white students in CPS from our multinomial regression beta estimates and found that



	Fall	Winter	Spring
Black (Male)	183.8	189.8	197.8
Black (Female)	185.6	193.0	198.6
Hispanic (Male)	187.1	193.2	201.1
Hispanic (Female)	187.3	193.0	200.2
Other (Male)	188.7	195.9	202.7
Other (Female)	190.9	196.1	201.6
White (Male)	191.7	197.3	206.1
White (Female)	191.1	196.1	204.5
Asian (Male)	197.2	203.9	209.9
Asian (Female)	194.8	201.1	206.8

FIGURE 1 Average Math performance of 2nd graders in Chicago among races and genders. *Note*: *Indicates ($p < .05$) from white male as reference category after Bonferroni correction; **Indicates ($p < .01$); ***($p < .001$); $^{+}$Indicates ($p < .05$) between genders within a racial category; $^{++}$Indicates ($p < .01$); $^{+++}$($p < .001$)

Mathematics disparities from unmet support needs did not change throughout the year.

This analysis encouraged us to engage these data more and think about what students within a broader lens may have their support needs met in relation to this original comparison (i.e., moving beyond only comparing race, or gender, in math achievement, which we argue is needed). Because we could not find any *race by gender disaggregated* estimates, nor regression stratifications that explore elementary Mathematics performance at different achievement levels, this led us to explore the race by gender interactions from our multinomial regression, as well as conduct a Median Split, which, then, showcased the emergence of students who started below the Median having mostly insignificant differences to white males, and no gender parity, while those starting above the Median had many significant differences. Through this we purpose our work beyond "gap-gazing" given that our approach was to emphasize nuanced estimations not yet conducted.

We also wanted to emphasize the importance of articulating the difference between our approach and one that could be defined intersectionally. Namely, Kimberlé Crenshaw's notion of intersectionality during her keynote at the Southbank Centre in 2016 described how intersectionality is "less about overlapping identities, and not primarily about identity." Instead, Crenshaw explains, intersectionality is concerned with the "structures that make some identities the consequence of and vehicle for vulnerability," and contends that if we are to disrupt discrimination and violence head-on, we must engage with the contexts that have led to "the exclusion of some people but not others" (Crenshaw, 2016). Intersectionality is inappropriate for our study, then, as we do not have the rich and thick data sources to think about the counter-storytelling/disruption central to investigating representational, political, and structural injustice.

In a similar vein, we also did not analyze student-aligned data to shed light on patterns across time where these historically marginalized groups could be disaggregated further beyond gender and their Median Split categorization. In sum, given that we did not take up intersectionality as the driving force for our research purpose, design of our methodology, and analytic schema, our inclusion of this term would invariably fall under the category of superficial use. To reiterate, we align with intersectional scholars when they describe the appropriate use of this term: "[An intersectional research inquiry is not defined primarily by using] the term "intersectionality," nor its being situated in a familiar genealogy, nor its drawing on lists of standard citations. Rather, what makes an analysis intersectional ... [is] thinking about the problem with sameness and difference and its relation to power" (Cho et al., 2013, p.795).

Given that our particular approach (race by gender, and then by Median Split) has not yet been conducted and published in relation to Mathematics performance for elementary students, we believed that such an analysis would help those who continue to speak with and through "achievement gap" and "gap-gazing" language to become more sensitive to the ways in which they could disaggregate both race by gender, and then include a regression stratification of estimated marginal means to think more complexly about within-group variance, and nuanced gradients of performance, which is one of the central arguments made by Gutiérrez (2008).

4 | FINDINGS

4.1 | Standards connections: Research Question 1

One of the main goals of this work was to understand opportunities in the Chicago's K-2 Mathematics standards to integrate spatial reasoning. We examined the K-2 Mathematics standards commonly used by district teachers (The Common Core Math Standards) for such opportunities. We used two categories to evaluate the extent to which spatial reasoning was already present in these standards and would be good candidates for the integration of spatial skills: *Strong* and *Weak*. Strong recognizes prior work in the research that indicates spatial reasoning is key to that Math topic. Weak means that there is no evidence from the literature that spatial reasoning would have an impact when integrated within this particular content of Mathematics. In Table 1 below, we show the findings from that analysis with shaded rows representing areas where we recommend infusing spatial reasoning; and unshaded not recommended. Two topics among the Common Core State Standards in Math rose prominently as high impact components that would benefit from the integration of spatial reasoning: Measurement and Data; and Geometry. These findings, then, informed our analysis of MAP Math Achievement among 2nd grade students in Chicago.

4.2 | Overall results of math achievement analyses: Research Question 2

To explore if any differences in Mathematics performance exist among 2nd graders, we conducted a Multinomial Logistic Regression and then calculated adjusted means for racial and gender comparisons. We sought to test if 2nd grade Math performance on the Measures of Academic Progress (MAP) significantly differed overall and among

TABLE 1 Common Core State Standards in Mathematics: Opportunities for spatial integrations

Common Core Math content focus	Grade level of curriculum	Percentage of curriculum with opportunities for spatial reasoning integration
Algebra	K	80%
	1st	25%
	2nd	25%
Base 10	K	0%
	1st	50%
	2nd	11%
Measurement and data	K	100%
	1st	100%
	2nd	100%
Geometry	K	100%
	1st	100%
	2nd	100%

We show the findings from that analysis with shaded rows representing areas where we recommend infusing spatial reasoning; and unshaded not recommended.

students after accounting for different school characteristics (i.e., percentage of students that receive free and reduced lunch [FRL]; percentage of students that are labeled as English Language Learners [ELLs]). These findings around early Math learning represent a sharpened set of analyses that can be leveraged to specifically target students who are differentially performing in elementary Mathematics and then test novel spatial thinking impact programs to ameliorate any academic achievement gaps by design. Overall, our analyses suggest that racial and gender disparities exist as early as the beginning of 2nd grade, and persist; additionally, students that start above the Median in 2nd grade exhibit these disparities disproportionately. These findings suggest a nuanced picture not yet explored by the field.

4.3 | Finding 1: At second grade, racial disparities exist in math performance, and persist

Our first analysis examined the aggregate MAP scores across all four Mathematics topics. In Tables 2 and 3 below, we provide our descriptive population statistics and likelihood ratio tests from the Multinomial Logistic Regression analyses. These results reveal differences in MAP math performance that can be predicted by students' race and that these differences not only exist at the beginning of grade 2 but persist throughout the year. The gender variable in this model indicates that there were no significant differences between the proportion of males and females for white students compared to any other racial group at each time point. Additionally, when including school-factors, these disparities still exist and persist, which provides further empirical support that each racial group is not being served equitably, no matter the school they attend. Figure 1 provides a graphical representation of these estimated means. Racial groups are greyscale coordinated across white, Black, Hispanic, and Other racial categories, with female groups within each race outlined in white denoting their means in comparison to their male counterparts. All significant comparisons are to white male students, as our multinomial regression results suggest white students outperform Black, Hispanic, and Other groups, and research has shown that males are often predicted to perform better on various math tasks than their female counterparts (Carmichael et al., 2014; Lowrie & Diezmann, 2011).

Unpacking an example of these results, in the Fall, Black and Hispanic racial groups (across both genders) showcase significantly less access to academic supports that could respond to their socio-cultural learning needs than their white male counterparts on this overall MAP assessment (denoted by three asterisks; $p < .001$). Additionally, Black female students, while still not being served appropriately compared to their white male counterparts, had significantly higher likelihood to have their academic support needs met as female students than their Black male counterparts (denoted by three vertically aligned plus signs [+++]; $p < .001$). Conversely, Asian students (both male and female) have benefited most from the type of Mathematics curriculum and instruction provided by Chicago Public Schools exhibited by significantly higher performance than white males. This pattern of disparity is sustained across the school year, with an additional disparity emerging between white male students

TABLE 2 Descriptive statistics and likelihood ratio tests from multinomial logistic regression models

Population samples (*N*)	Time 1		Time 2		Time 3	
Total	14,772 [174.1 (14.6)]		21,360 [181.3 (13.9)]		24,216 [190.8 (14.0)]	
Asian	683 [185.3 (16.0)]		950 [193.3 (14.4)]		1414 [199.6 (14.2)]	
Black	6274 [170.4 (13.7)]		8102 [177.6 (13.6)]		8438 [187.2 (14.2)]	
Hispanic	5607 [173.6 (13.2)]		8513 [180.5 (12.3)]		9978 [189.7 (12.5)]	
Other	533 [178.8 (15.1)]		1254 [183.4 (13.2)]		1338 [192.3 (14.0)]	
White	1675 [183.7 (14.2)]		2541 [190.4 (13.0)]		3048 [199.8 (12.6)]	
Female	7321 [174.4 (13.8)]		10,538 [181.3 (13.1)]		11,970 [190.5 (13.4)]	
Male	7451 [173.9 (15.4)]		10,822 [181.4 (14.6)]		12,246 [191.2 (14.6)]	
Asian males	378 [186.5 (16.5)]		497 [194.7 (15.3)]		734 [201.1 (14.9)]	
Asian females	305 [183.8 (15.2)]		453 [191.9 (13.1)]		680 [197.9 (13.2)]	
Black males	3140 [169.5 (14.2)]		4049 [176.9 (14.0)]		4221 [186.8 (14.6)]	
Black females	3134 [171.3 (13.2)]		4053 [178.3 (13.2)]		4217 [187.6 (13.8)]	
Hispanic males	2798 [173.6 (13.8)]		4301 [180.6 (12.9)]		5026 [190.2 (13.0)]	
Hispanic females	2809 [173.7 (12.6)]		4212 [180.4 (11.5)]		4952 [189.2 (12.0)]	
Other males	274 [178.0 (16.6)]		643 [184.4 (13.9)]		685 [193.0 (14.5)]	
Other females	259 [179.7 (13.3)]		611 [183.2 (12.4)]		653 [191.7 (13.3)]	
White males	861 [184.0 (14.8)]		1332 [191.0 (13.6)]		1580 [200.6 (12.8)]	
White females	814 [183.5 (13.5)]		1209 [189.8 (12.4)]		1468 [199.0 (12.3)]	
	Time 1 (no school)	**Time 1 (with school)**	**Time 2 (no school)**	**Time 2 (with school)**	**Time 3 (no school)**	**Time 3 (with school)**
R-squared						
Nagelkerke (*McFadden*)	0.119 (*0.047*)	0.653 (*0.367*)	0.129 (*0.049*)	0.679 (*0.382*)	0.112 (*0.042*)	0.633 (*0.336*)
Independent variables	Likelihood ratio tests: Chi-square test estimates					
Gender	7.082	4.494	2.437	2.242	0.433	0.582
Math score	1702.7***	464.1***	2715.0***	718.9***	2664.0***	830.5***
FRL % at school	—	3122.2***	—	5611.5***	—	5989.2***
ELL % at school	—	9203.1***	—	13,965.1***	—	13,853.4***

Note: Notation: Sample size [mean (std. dev.)].

***$p < .001$.

and those students in the "Other" racial groups (e.g., Native Hawaiian; Native American; Multi-racial; and Pacific Islander).

By Spring, the gender disparity between male and female Black students is no longer significant, but Hispanic, white, and Asian female students emerge not having their academic support needs met on this MAP assessment in terms of elementary Mathematics instruction in Chicago. In sum, Figure 1 indicates that the racial disparities in MAP Mathematics scores exist even as students come into 2nd grade and that gender disparities emerge by the end of 2nd grade. Taken further, in Figure 2 we graphically represent the bias-corrected effect sizes between all racial comparisons to white males. These effect sizes, rather than merely describing significant differences, provide a measure of impact between our comparison group (white males) that is also adjusted based on differences in variances that may exist due to sample size differences.

What Figure 2 illuminates is similar to the narrative above: These race and gender disparities do not just sustain at the same level of difference, they get worse. We again see the emergence of Other students also being unsupported academically across 2nd grade. Across these time points, though, the disparities in MAP math performance between Asian and white students are less prevalent, and in the case of white females there is no effect by the end of the year. This suggests that the curriculum and pedagogies employed in Chicago public schools serve white and Asian students most and may not be well-designed to ameliorate racial disparities at this level. And, to reiterate, these mean scores for all groups are adjusted such that the percentage of students who qualify for Free and Reduced Lunch and who are labeled as English-language Learners are equal

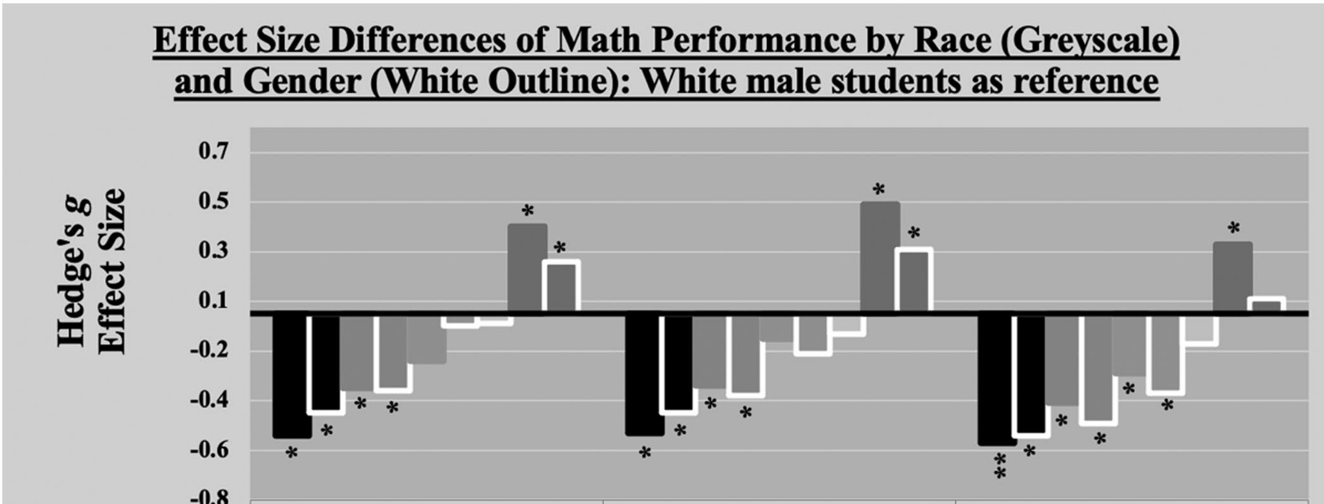
TABLE 3 Parameter estimates: Multinomial logistic regression models across race for all 2nd grade Math goals

Race[a]	Independent variables	Time 1 (no school factors)	Time 1 (with school factors)	Time 2 (no school factors)	Time 2 (with school factors)	Time 3 (no school factors)	Time 3 (with school factors)
Asian	Gender	−0.148 (0.091)	−0.150 (0.094)	0.039 (0.077)	0.064 (0.079)	−0.006 (0.065)	0.022 (0.068)
	Math performance	0.008* (0.003)	0.028*** (0.004)	0.019*** (0.003)	0.040*** (0.003)	−0.002 (0.003)	0.017*** (0.003)
	Free and reduced lunch % at school (FRL %)	—	1.528*** (0.268)	—	1.847*** (0.224)	—	4.824*** (0.168)
	English language learner % at school (ELL %)	—	2.373*** (0.368)	—	1.544*** (0.316)	—	−3.288*** (0.232)
Black	Gender	0.059 (0.057)	0.027 (0.073)	0.062 (0.048)	0.087 (0.065)	−0.001 (0.044)	0.031 (0.058)
	Math performance	−0.072*** (0.002)	−0.036*** (0.003)	−0.080*** (0.002)	−0.039*** (0.003)	−0.075*** (0.002)	−0.045*** (0.002)
	FRL %	—	8.007*** (0.190)	—	8.916*** (0.171)	—	8.447*** (0.151)
	ELL %	—	−11.477*** (0.286)	—	−12.082*** (0.249)	—	−11.154*** (0.213)
Hispanic	Gender	0.047 (0.057)	0.020 (0.064)	0.032 (0.047)	0.043 (0.055)	−0.012 (0.043)	0.009 (0.050)
	Math performance	−0.055*** (0.002)	−0.027*** (0.002)	−0.062*** (0.002)	−0.029*** (0.002)	−0.062*** (0.002)	−0.033*** (0.002)
	FRL %	—	4.753*** (0.176)	—	5.665*** (0.153)	—	5.652*** (0.138)
	ELL %	—	0.274 (0.246)	—	−0.762*** (0.214)	—	−0.879*** (0.183)
Other	Gender	−0.022 (0.100)	−0.028 (0.103)	0.002 (0.070)	0.014 (0.074)	−0.035 (0.066)	−0.015 (0.070)
	Math performance	−0.027*** (0.004)	−0.012** (0.004)	−0.043*** (0.003)	−0.010** (0.003)	−0.046*** (0.003)	−0.023*** (0.003)
	FRL %	—	4.688*** (0.230)	—	2.289*** (0.218)	—	2.574*** (0.196)
	ELL %	—	−7.944*** (0.396)	—	3.004*** (0.282)	—	1.651*** (0.247)

Note: Notation: Beta (std. error).

[a]Reference category are white students.

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.



	Fall	Winter	Spring
Black (Male)	-0.49	-0.48	-0.52
Black (Female)	-0.40	-0.40	-0.49
Hispanic (Male)	-0.30	-0.29	-0.36
Hispanic (Female)	-0.31	-0.33	-0.44
Other (Male)	-0.19	-0.10	-0.24
Other (Female)	-0.05	-0.16	-0.32
White (Female)	-0.04	-0.08	-0.12
Asian (Male)	0.35	0.44	0.28
Asian (Female)	0.21	0.26	0.06

FIGURE 2 Hedge's g bias-corrected effect sizes of average Math performance of 2nd graders. *Note*: *Indicates small effect sizes ($0.2 > g < 0.49$); **Moderate effect sizes ($0.5 > g < 0.79$)

across all groups, which suggests that even if these students attended schools with similar socio-economic status and language fluency make-up, we would likely predict that these patterns of student disparity would still emerge and persist at this grade level.

4.4 | Finding 2: MAP math comparisons of below the median starting 2nd graders

Our second and third more detailed analyses sought to compare both differences in racial categories and genders for the MAP math assessment, because we had found that racial disparity not only exists at 2nd grade math performance but persists across the grade level (see above). Given that we identified two out of the four goals within the MAP math assessment that would be viable candidates to integrate a spatial reasoning intervention (see section above), we explored if any difference existed among these two goals: (c) Measurement and Data and (d) Geometry. These Estimated Marginal Means provided scores of each of these sub-group interactions among race and gender (e.g., white males compared to Black females; white males compared to Hispanic males) after adjusting for the percentage of free and reduced lunch students in a school and the percentage of English-language learners in a school—similar to Finding 1's modeling. We additionally disaggregated race by gender student groups across their starting achievement level for Fall at the median, adjusting that median value based on white male growth patterns. Again, this referent of white males allows for discussions to be had about the persistence of curriculum and instruction in public schooling to center their academic support needs most.

For Finding 2, we explored below the median 2nd graders at each time point and calculated any significant effect sizes that are also corrected for bias based on differences in standard deviations between groups. In Figure 3, each race is represented by the same greyscale used in Figures 1 and 2; female genders within each race are outlined in white directly to the right of males. At the base of a particular column there is a white number (i.e., the bias-corrected effect size) if that particular student group is significantly different than white males at that time point (e.g., for time point 2 [Winter], Black males Mathematics academic support needs were less supported than their white male counterparts, and the magnitude of

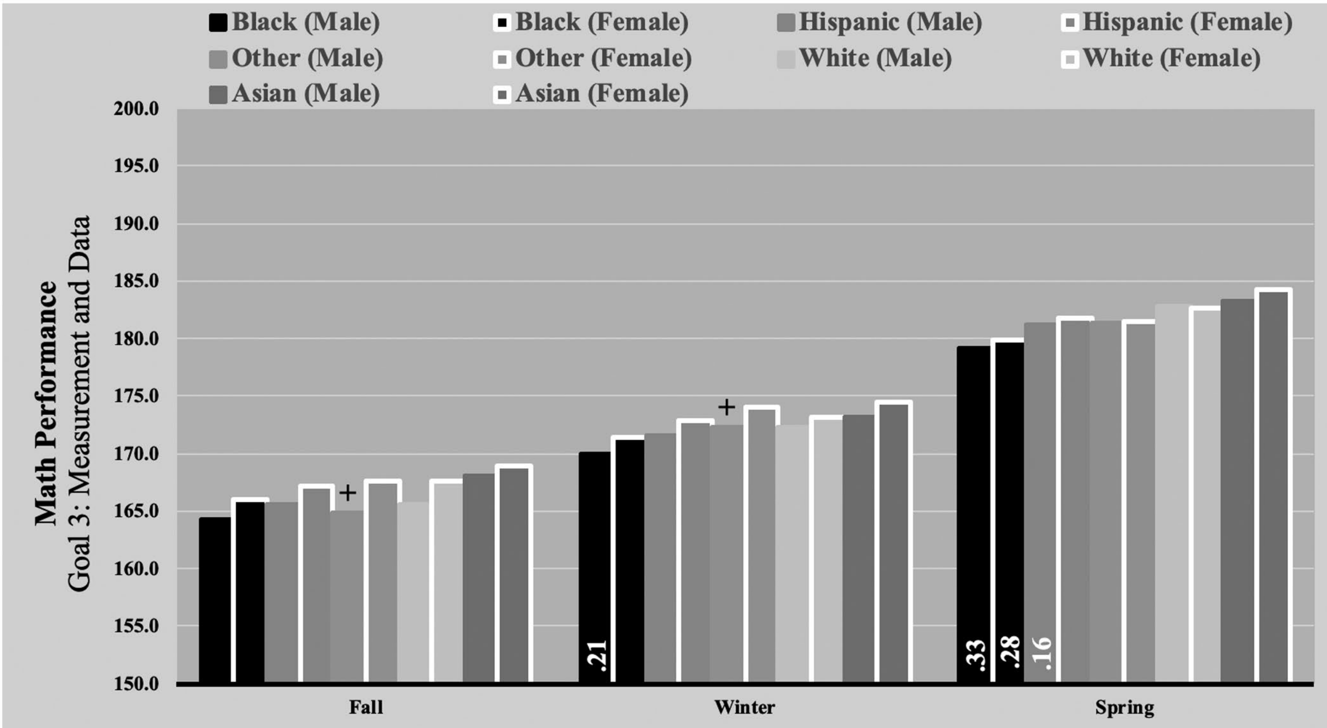


FIGURE 3 Below the median student differences for Goal C (NWEA Goal 3: Numbers and measurement)

that difference is 0.21). Also, there are (+) signs above the male column of a race if there is a significant difference and small effect size between males and females within a race (e.g., for time 2, Other males did not have their support needs met compared to their female counterparts, at a small effect size).

What we find for below the median student differences for Goal C (Measurement & Data) in Figure 3 is that as we move from Fall (where there are no differences in whose support needs are met between race by gender categories and their white male counterparts) to Spring (where there are multiple groups of students who are not having their socio-cultural support needs met when compared to white males). In sum, Black students (male and female), as well as Hispanic males, were not being served well when compared to their white male counterparts—although, no other races exhibit this MAP Math performance disparity. And while we see gender differences within the Other racial category in Fall and Winter, this is not present by Spring, alluding to male Other students becoming more supported in their academic support needs in Chicago Public Schools compared to their female Other counterparts.

Below in Figure 4, we also see no significant gender differences among any race at any time point for Goal D (Geometry); moreover, the only race not having their academic support needs met other than white males are Black males, and Black females emerge as not being supported by the Spring time point. This, combined with our Finding 1 analyzing racial disparities aggregated across genders, suggests that while Black males may start more likely to score lower on the MAP math assessment in 2nd grade, Black females by Spring are also predicted to additionally not have their assets valued in terms of the 2nd grade Mathematics instruction when compared to their white male counterparts. In sum, between Figures 3 and 4, our Finding 2 suggests that students that may be characterized as below the median exhibit the prominent achievement gap and racial disparity that was found in Finding 1 only for Black students and emerge by the end of 2nd grade among Hispanic male students for Goal C. This analysis, thus, sheds light on the reality that different stratifications of achievement across race and gender should examine how to approach these students' nuanced support needs. These analyses also suggest that while we might perceive below the median students as the source of where programs around spatial reasoning would be most impactful to ameliorate racial and gender disparities related to achievement gaps, that assumption is not supported by these data.

4.5 | Finding 3: MAP math comparisons of above the median starting 2nd graders

As stated above, our second and third findings represent analyses of above and below the median students

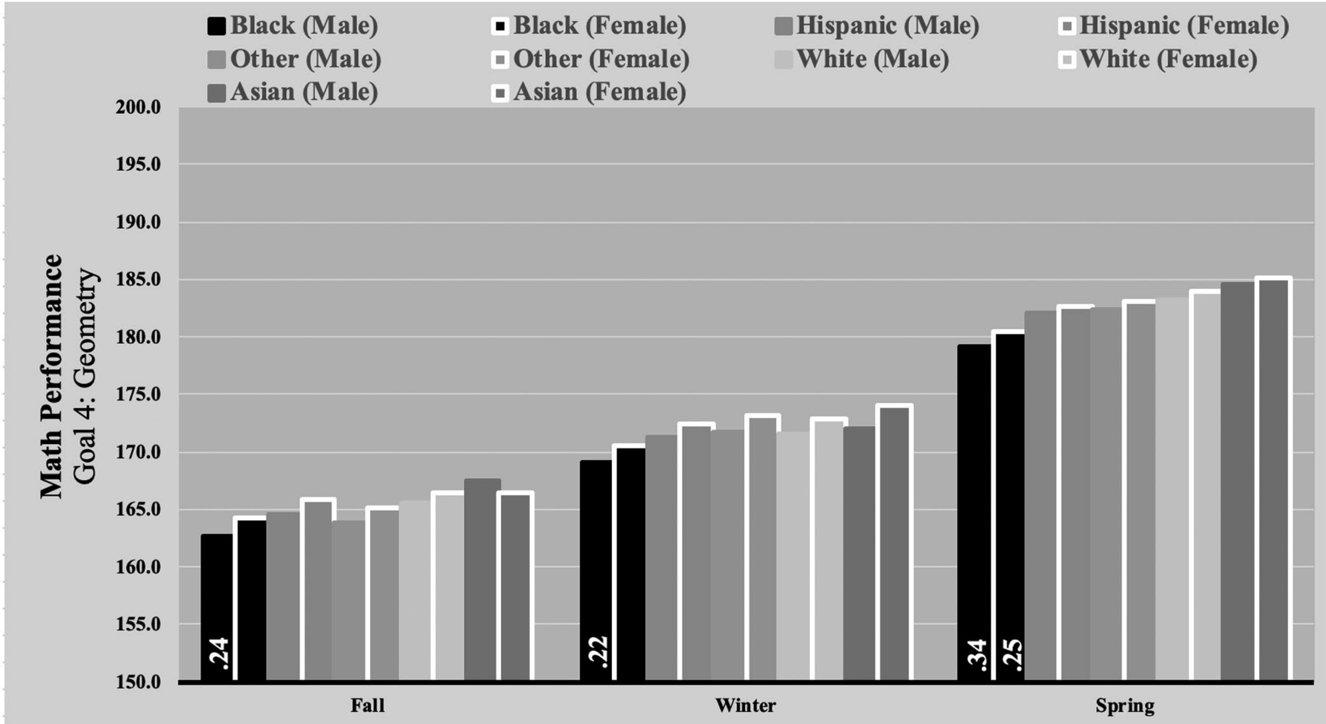


FIGURE 4 Below the median student differences for Goal D (NWEA Goal 4: Geometry)

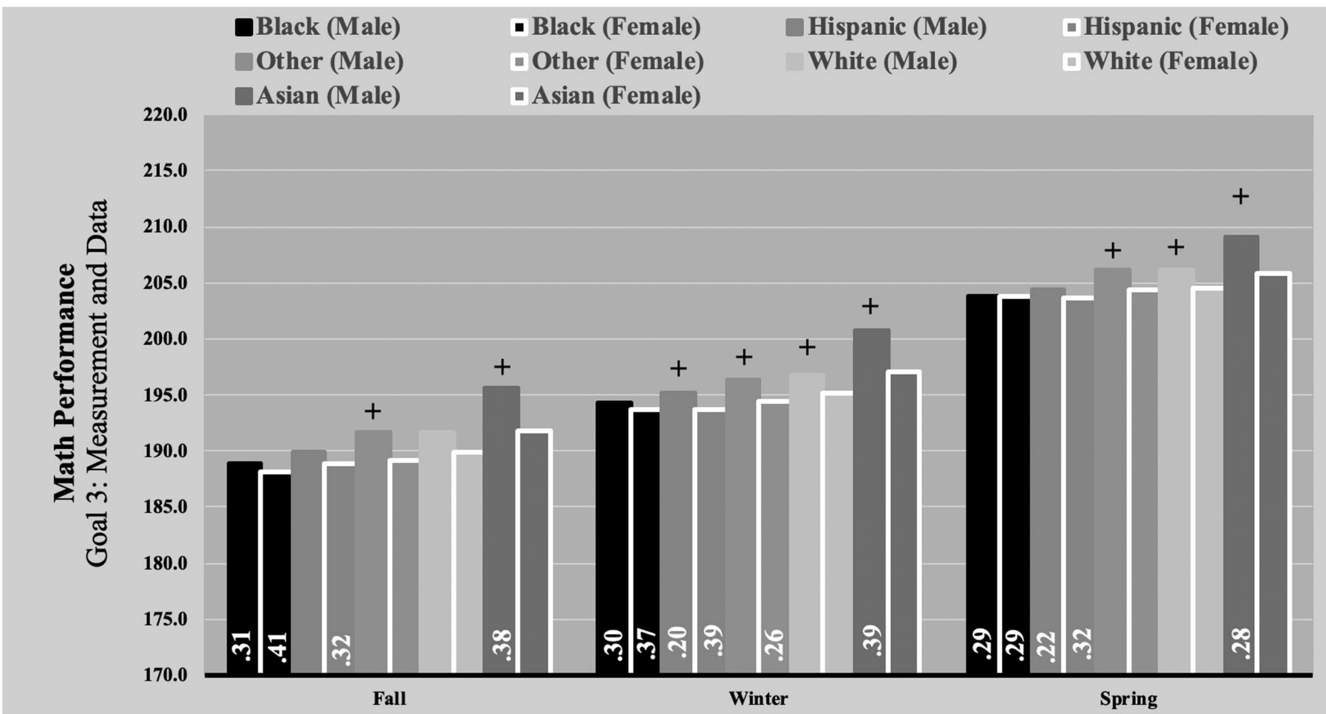


FIGURE 5 Above the median student differences for Goal C (NWEA Goal 3: Numbers and measurement)

categorized by the Fall median across the entire data set and then adjusting that median value relative to the growth of white males in the data set. Utilizing the same notations as the above Figures 3 and 4, below we explore the above the median stratification of 2nd grade students across MAP math goals C and D—those goals we have identified as leverage points where we can envision infusions of spatial reasoning programs to ameliorate the achievement gap. First, Figure 5 below showcases the comparisons among races and genders for Goal C, while Figure 6 below showcases

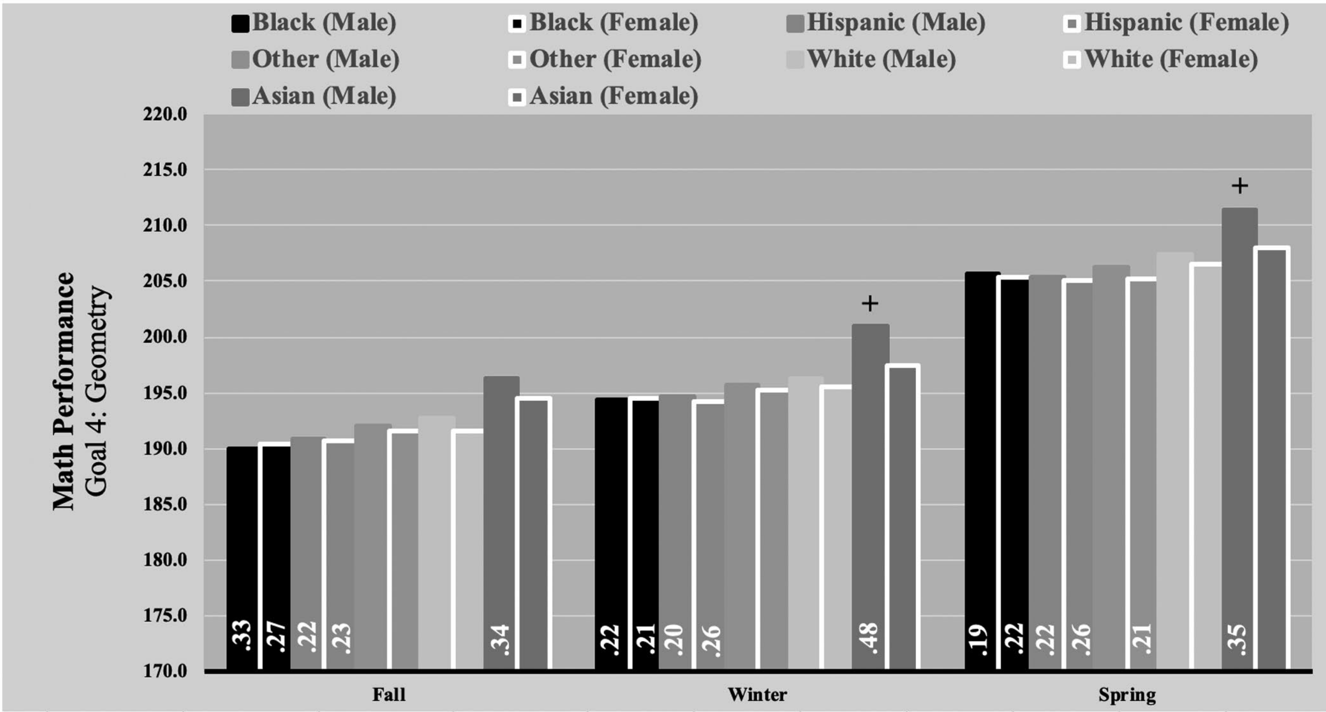


FIGURE 6 Above the median student differences for Goal D (NWEA Goal 4: Geometry)

these same comparisons for Goal D. Each of these Figures are described in detail for their important highlights related to racial disparity, as well as any gender disparities that may exist within racial social categories of students.

In Figure 5, compared to Figures 3 and 4, we see a substantial difference in the prevalence of differential performance on the MAP Math assessment for Goal C (Measurement & Data), both around racial and gender disparities. To start, like Figures 3 and 4 in Finding 2, this profile analysis showcases how Black students of both genders are consistently not supported in their Mathematics learning when compared to their white male counterparts. This is the same for Hispanic male and females, as well. However, while in Figures 3 and 4 we found an increase of the effect size (i.e., the magnitude of difference between white males and the group being compared; in white numbers at the base of the columns), here these differences seem to lessen for Black males and females, albeit for Black males this dampening of difference in support from 0.31 to 0.29 is much less of a dampening of the achievement gap for Goal C than their Black female counterparts who show a substantial decrease in this support need gap from 0.41 to 0.29. It should be noted that for Hispanic males this academic support gap above the median appears to be widening from no difference in Fall to 0.22 in the Spring, while Hispanic female academic support needs in Chicago remain consistently, albeit unwavering, lower than their white male counterparts (~0.32). This teases out a nuanced picture for Mathematics disparity.

Pertinent to our goals in identifying where spatial integrations could impact disparities, Figure 5 shows the most gender disparities within racial categories. While in the Fall, only Other male and Asian males showcase greater support compared to their female counterparts, in the Winter time point we see an emergence of lacking support on Goal C in the MAP math assessment for Hispanic and white females, as well. And while Hispanic females within the Spring time point do not exhibit differential support compared to their Hispanic male counterparts, white females maintain their disparate support in 2nd grade Mathematics compared to their white male counterparts. Additionally, across all three time points females are predicted to not have their socio-cultural support needs met across most races. These disparities among race and gender illuminate a starkly different picture than those in the below the median category we presented in Figure 3. In fact, in general, females in the below the median categorization for Goal C tended to have higher scores on this MAP math goal than their male counterparts, and even significantly higher at one time and race: Other female, Winter. This suggests a prime area of interest to inquire about the possibility of implementing programs to improve spatial reasoning among race and gender related to Goal C on the MAP assessment, which aligns well with the findings from Goal D among this above the median category, unpacked below.

Figure 6 showcases similar racial disparities for Goal D as the below the median students we have showcased in Figure 4. Namely, Black males and females show persistent differential support needs (un)met for Goal D. Intriguingly, within the below the median stratification, Black male effect sizes in Figure 4 showcase an increase from 0.24 to 0.34 across 2nd grade for Goal D, while Figure 6 showcasing above the median stratification exhibits Black male effect sizes decreasing to the point of negligible effects, from 0.33 to 0.19, when compared to their white male counterparts. Black females in the above the median stratification group showcased in Figure 6 showcase a similar decrease of disparity, from 0.27 to 0.22, alluding to Black students' support needs being met in relation to within-race support across gender for Goal D (Geometry), but only for above the median students. In turn, we also find that above the median Hispanic students in Figure 6 persist in Math disparities across genders alluding to a relatively stable and unmediated unmet support needs when compared to their white male counterparts (0.22–0.26, as a range), which seems to be the same case for above the median Asian male students (0.34–0.35).

However, no gender differences in MAP math performance are significant other than Asian males having their academic and socio-cultural support needs met better than their female counterparts. It also seems that the same general trend found in Goal C when comparing above and below the median stratified gender disparities (albeit more subtle in Goal D), is also found in Goal D; specifically, below the median females tend to have their support needs met better than their male counterparts while above the median females exhibit less support than their male counterparts. Taken together these Figures showcase significant and substantial trends in MAP Math performance across Goals C and D—those Goals we have identified as pertinent leverage points for spatial reasoning program inclusions—that can provide insight into ways to design more equitable learning opportunities for all learners across race, gender, achievement, and their overlapping nexus. Through this analytic technique, we contribute to the field's approaches among disparity analyses to illuminate nuanced approaches to race, gender, and achievement Mathematics analyses, and in particular pointing to the need for stratified explorations of who is having their socio-cultural and academic support needs met in Chicago Public Schools.

5 | DISCUSSION

In this research we (1) make the case for integrating spatial reasoning into existing elementary Mathematics curriculum areas to improve all students' performance and inevitably ameliorate disparities by design that emerge as a function of unmet socio-cultural and academic supports; (2) provide alignments between these spatial reasoning integrations that could be specifically designed to improve the Common Core State Standards that are widely used; (3) analyze the third largest urban school district in the US in terms of 2nd grade Mathematics performance across race; and (4) sharpen the approach to analyzing race and gender interactions that exist among the racial disparities we found in relation to different achievement levels of students. Our work aligns with previous research reports that consistently find Mathematics disparities that emerge in elementary grades (Ottmar et al., 2013; Pianta et al., 2008), and contributes to work that provides support for specific programs to ameliorate such racial and gender disparities (Brand et al., 2006; Ford, 1994; Nicol & Crespo, 2005; Sparks & Pole, 2019).

Our greatest contribution to the field of urban elementary Mathematics is our analyses that illuminated where disparities do and do not exist. We predicted that if racial and gender disparities existed, we would see these differences at all Math achievement levels, but this was not the case. Namely, we expected all students who were above the median when they started 2nd grade may (a) be on par with their white male counterparts, or (b) would be able to catch up due to their achievement starting point. Instead, Figures 3 and 4 suggest Black and Hispanic males and Black females below and above the median are those student groups that are not having their socio-cultural support needs met. All other race by gender interaction groups were not significantly different from white males, as well as no within-group racial differences based on gender were significant. Conversely, for those students that started above the median, both Black and Hispanic males and females are not being supported when compared to white males, and there are significant gender differences within white and Asian student racial groups (see Figures 5 and 6). This calls into question how we conceptualize who needs support, and what programs should be designed into any existing standards, curriculum, and pedagogy, as well as how we should think about analyzing Mathematics performance disparities.

One such example, we feel, makes a strong case for our argument. In our analysis, it is evident that Black and Hispanic male students are not being served appropriately in elementary Mathematics across all achievement levels, and that Black females emerge as not being supported for their socio-cultural needs in Mathematics by the end of 2nd grade. However, students starting above the median showcase significant gender disparity that is not present for students that start second grade below the median. This means that if we want to approach elementary Mathematics reform, we need to support Black and Hispanic males in ways that draw on their assets, and support females who come into 2nd grade proficient in math. Moreover, we need to draw on the meta-analytic and impact studies we unpacked above to co-design these interventions with teachers (cf. Lowrie et al., 2017; Yang et al., 2020). We provide some suggestions below as to how to integrate spatial reasoning among the CCSS with teachers.

5.1 | Recommendations on spatial reasoning integration for the CCSS at K-2 grades

We recommend focusing on two disciplinary Mathematics practices easily adopted in existing curriculum and pedagogy. These practices were chosen given that they both have a strong potential to impact students' spatial reasoning (see background) that are crucial for the two CCSS in Math around Goal C (Measurement & Data) and Goal D (Geometry): i.e., (a) Data Analysis and (b) Modeling. We recommend integrating the following types of tasks into data analysis and modeling instruction among the previous two CCSS in Mathematics: (a) Creating data representations (e.g., maps, diagrams, and graphs); and (b) Constructing models (e.g., physical models, explanatory models). Research suggests that repeated practice with such tasks can improve spatial reasoning and impact Math performance (Uttal, 2000; Newcombe, 2010).

For data analysis tasks, students can create visual representations of data they collect during student-centered learning activities. These activities can involve a variety of graphs and charts, such as bubble charts, topological and heat maps, and frequency graphs of count data. During these tasks, students should be asked to reason how ways of representing data are different and how these representations impact what they notice about the data. Students should also be supported to observe patterns in data. Two types of models that we recommend focusing on during modeling tasks are explanatory models and physical models. To create explanatory models, students sketch the components of a phenomenon or concept they are learning and use lines, arrows, and other shapes to illustrate their relationships. For example, students could draw explanatory models that illustrate how input values for a simple equation lead to predictable outputs. Students should also be asked to reason about the models they created, including the choices they made in creating them, how the different parts of the process relate to one another, and how that process leads to the outcome. Students can also construct and engage with a variety of physical models. For example, students can use blocks, clay, and other manipulatives to create models of shapes and patterns. This can progress from simple shapes, like triangles and rectangles, to analyzing more complex Mathematical patterns, such as symmetry. Students can also use folded paper or Lego blocks to create more complex patterns. Block building tasks of these types have been shown to impact student's spatial reasoning skills (Casey et al., 2008).

Across all these tasks, students are engaged in not only creating graphs, models, and representations but also making sense of their patterns and relationships, as well as thinking about what and how they learned. When making sense of data representations, students should, additionally, reflect on similarities and differences among forms of representation and their affordances for sense making. Similarly, they should reflect on the models they create such as how and why they made particular representational choices and how and why the different parts of their model relate to one another. Spatial language and talk that describes locations and relationships in space, and gestures, can also help students learn to think spatially (Sauter et al., 2012; Goldin-Meadow & Singer, 2003; Newcombe, 2010; Simms & Gentner, 2009). We also suggest that teachers frequently use spatial words, such as between, above, outside, under, and around, and indicate spatial orientations and relationships through gesture during these tasks. In turn, integrating such changes to support spatial reasoning in elementary Mathematics also sheds light on the powerful implications of our work in relation to supporting all students' learning.

6 | IMPLICATIONS

Across our analyses, the data support that below the median students do not widely differ across 2nd grade in Mathematics, even as that would be the area where many would expect needed programs of support. In turn, we draw on work like Hill et al. (2008) who have pointed out that programs to ameliorate achievement gaps in math showcase differential impacts related to how proximal the program is to the measurement variable. In other words, the MAP math performance assessment, while not specific to spatial reasoning, does allude to an alignment between the affordances that spatial improvement can provide with population support. We argue that drawing on elementary Math spatial ability programs that have shown significant effect sizes such as Lowrie et al. (2017; $d = 0.4$) who have successfully been implemented to improve such spatial reasoning is crucial for equitable future endeavors in elementary Mathematics. Moreover, as Rutherford et al. (2014) report in their disaggregated program for elementary Mathematics spatial ability, there has been shown negligible impacts for lower-level Math ability students but small effects for moderate to higher-level students in their study (d = 0.16–0.20). Given our analysis suggests dampened racial and gender disparity at below the median levels, this work refocuses our pursuits toward equitable Math achievement for all students specific to designing explicitly to meet students' socio-cultural and academic support needs.

With the onset and integration of the Common Core State Standards in Chicago Public Schools and throughout the country, our work provides both a statistical view of where racial and gendered disparities in elementary Mathematics may emerge and persist. However, we do not leave the argument there; instead, we draw from work on spatial reasoning to suggest specific areas within the CCSS to integrate a research-based approach to addressing and ameliorating the disparities we found among our student populations. We also challenge current work done around understanding the implementation of the Common Core States Standards in Mathematics among elementary contexts (i.e., Schweig et al., 2020) in that the standards provide no guidance on the need for spatial reasoning and that this type of integration of such an important Mathematics topic is not overlooked but, in fact, invisible (National Research Council, 2006).

The implications of the research we have presented suggest that while there may be more systematic impacts due to the disproportionate number of students that qualify for Free and Reduced Lunch at certain schools, these socio-economic impacts do not fully account for the variance in achievement among racial and gendered elementary students. Invariably, we align with broader policy suggestions around the role of administration in the process of ambitious elementary Mathematics instruction, as well, in that the guidance for teachers to engage in curricular design and implementation must have productive and generative support from their principals (Rigby et al., 2021). Finally, through our statistical analyses and careful coding of areas in the CCSS that would benefit from the rich research-based impacts of spatial reasoning, we provide not just an argument of lack but, rather, also hopeful solutions that could be designed toward equitable Mathematics for all, as well as an approach for researchers focused on disparity in Mathematics to sharpen their focus to within-group variability in relation to historically marginalized students' socio-cultural and academic support needs (Gutiérrez, 2008). Through this work our data suggest that while we must examine racism embodied by the normative center of schools, we must also attend to the ways female students support needs are met equitably.

ORCID

Phillip A. Boda https://orcid.org/0000-0001-5797-8139

REFERENCES

Bachman, H. J., Votruba-Drzal, E., El Nokali, N. E., & Castle Heatly, M. (2015). Opportunities for learning math in elementary school: Implications for SES disparities in procedural and conceptual math skills. *American Educational Research Journal*, *52*, 894–923. https://doi.org/10.3102/0002831215594877

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology*, *53*(1), 371–399.

Brand, B. R., Glasson, G. E., & Green, A. M. (2006). Sociocultural factors influencing students' learning in science and mathematics: An analysis of the perspectives of African American students. *School Science and Mathematics*, *106*, 228–236. https://doi.org/10.1111/j.1949-8594.2006.tb18081.x

Brisson, B. M., Dicke, A. L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal*, *54*, 1048–1078. https://doi.org/10.3102/0002831217716084

Cameron, C. E., Grimm, K. J., Steele, J. S., Castro-Schilo, L., & Grissmer, D. W. (2015). Nonlinear Gompertz curve models of achievement gaps in mathematics and reading. *Journal of Educational Psychology*, *107*, 789–804. https://doi.org/10.1037/edu0000009

Carmichael, C., MacDonald, A., & McFarland-Piazza, L. (2014). Predictors of numeracy performance in national testing programs: Insights from the longitudinal study of Australian children. *British Educational Research Journal*, *40*, 637–659. https://doi.org/10.1002/berj.3104

Casey, B., Andrews, N., Schindler, H., Kersh, J., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction*, *26*, 269–309. https://doi.org/10.1080/07370000802177177

Cho, S., Crenshaw, K. W., & McCall, L. (2013). Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society*, *38*, 785–810. https://doi.org/10.1086/669608

Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education*, 192–212. https://doi.org/10.2307/749610

Crenshaw, K. (2016, March 15). On intersectionality [Video]. Youtube. Retrieved from: https://www.youtube.com/watch?v=-DW4HLgYPlA&t=190s

Ford, M. I. (1994). Teachers' beliefs about mathematical problem solving in the elementary school. *School Science and Mathematics*, *94*, 314–322. https://doi.org/10.1111/j.1949-8594.1994.tb15683.x

Garcia-Olp, M., Nelson, C., & Saiz, L. (2019). Conceptualizing a Mathematics curriculum: Indigenous knowledge has always been Mathematics education. *Educational Studies*, *55*, 689–706. https://doi.org/10.1080/00131946.2019.1680374

Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology*, *55*, 637–652. https://doi.org/10.1037/dev0000649

Gilligan, K. A., Hodgkiss, A., Thomas, M. S., & Farran, E. K. (2019). The developmental relations between spatial cognition and mathematics in primary school children. *Developmental Science*, *22*(4), 1–19.

Goldin, C. D., & Katz, L. F. (2009). *The race between education and technology*. Harvard University Press.

Goldin-Meadow, S., & Singer, M. (2003). From children's hands to adults' ears: Gesture's role in the learning process. *Developmental Psychology*, *39*, 509–520. https://doi.org/10.1037/0012-1649.39.3.509

Gutiérrez, R. (2008). Research Commentary: A gap-gazing fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, *39*(4), 357–364.

Hadi-Tabassum, S. (2017). Navigating the STEM landscape: Examining the role of spatial reasoning for women of color. In J. Ballenger, B. Polnick, & B. Irby (Eds.), *Research on women and education. Women of color in STEM: Navigating the workforce* (pp. 135–155). IAP Information Age Publishing.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin and Review*, *27*(3), 465–482. https://doi.org/10.3758/s13423-019-01694-7

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education*, *4*(3), 60–68. https://doi.org/10.1016/j.tine.2015.05.001

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Holmes, J., Adams, J. W., & Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology*, *20*(2), 272–289.

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, *24*(9), 1831–1836. https://doi.org/10.1177/0956797613478615

Kuzmina, Y., & Ivanova, A. (2018). The effects of academic class composition on academic progress in elementary school for students with different levels of initial academic abilities. *Learning and Individual Differences*, *64*, 43–53. https://doi.org/10.1016/j.lindif.2018.04.004

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. https://doi.org/10.1037/bul0000191

Lee, J. (2010). Tripartite growth trajectories of reading and math achievement: Tracking national academic progress at primary, middle, and high school levels. *American Educational Research Journal*, *47*, 800–832. https://doi.org/10.3102/0002831210365009

Lee, J., & Wong, K. K. (2004). The accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, *41*(4), 797–832.

Leonardo, Z., & Broderick, A. (2011). Smartness as property: A critical exploration of intersections between whiteness and disability studies. *Teachers College Record*, *113*, 2206–2232. https://doi.org/10.1177/016146811111301008

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 127–155. https://doi.org/10.1002/wcs.1380

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 1479–1498. https://doi.org/10.2307/1130467

Lohman, D. F. (1993). *Effects of practice and training on the acquisition and transfer of spatial skills: Two speed-accuracy studies*. (Final Report Grant AFOSR-91- 9367). Lindquist Center for Measurement.

Lowrie, T., & Diezmann, C. M. (2011). Solving graphics tasks: Gender differences in middle-school students. *Learning and Instruction*, *21*, 109–125. https://doi.org/10.1016/j.learninstruc.2009.11.005

Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *British Journal of Educational Psychology*, *87*, 170–186. https://doi.org/10.1111/bjep.12142

Maeda, Y., & Yoon, S. Y. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue spatial visualization tests: Visualization of rotations (PSVT: R). *Educational Psychology Review*, *25*(1), 69–94. https://doi.org/10.1007/s10648-012-9215-x

Miles, M. L., Buenrostro, P. M., Marshall, S. A., Adams, M., & McGee, E. O. (2019). Cultivating racial solidarity among Mathematics education scholars of color to resist white supremacy. *The International Journal of Critical Pedagogy*, *10*, 97–125.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*, 1063–1066. https://doi.org/10.1111/j.1467-9280.2008.02200.x

National Research Council. (2006). *Systems for state science assessment. Committee on test design for K–12 science achievement*. Center for Education, Division of Behavioral and Social Sciences and Education. The National Academies Press.

Newcombe, N. S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *American Educator*, *34*(2), 29.

Newcombe, N. S. (2013). Seeing relationships: Using spatial thinking to teach science, mathematics, and social studies. *American Educator*, *37*(1), 26.

Ngo, F. J., & Velasquez, D. (2020). Inside the math trap: Chronic math tracking from high school to community college. *Urban Education*, (Online First). https://doi.org/10.1177/0042085920908912

Nicol, C., & Crespo, S. (2005). Exploring mathematics in imaginative places: Rethinking what counts as meaningful contexts for learning mathematics. *School Science and Mathematics*, *105*, 240–251. https://doi.org/10.1111/j.1949-8594.2005.tb18164.x

NWEA. (2019). *Linking study report: Predicting performance on the Illinois Assessment of Readiness (IAR) based on MAP® Growth™ scores*. Northwest Evaluation Association.

Ottmar, E. R., Konold, T. R., Berry, R. Q., Grissmer, D. W., & Cameron, C. E. (2013). Increasing equity and achievement in fifth grade mathematics: The contribution of content exposure. *School Science and Mathematics*, *113*, 345–355. https://doi.org/10.1111/ssm.12035

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*, 365–397. https://doi.org/10.3102/0002831207308230

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*, 1067–1070. https://doi.org/10.1111/j.1467-9280.2008.02201.x

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology*, *91*(2), 137–157.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*, 853–891. https://doi.org/10.3102/0002831209333184

Rigby, J. G., Forman, S., & Lewis, R. (2021). Principals' leadership moves to implement a discipline-specific instructional improvement policy. *Leadership and Policy in Schools*, *20*(2), 236–260. https://doi.org/10.1080/15700763.2019.1668422

Ross, J. A. (1995). Students explaining solutions in student-directed groups: Cooperative learning and reform in mathematics education. *School Science and Mathematics*, *95*, 411–416. https://doi.org/10.1111/j.1949-8594.1995.tb10194.x

Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., Richland, L., Tran, N., Schneider, S., Duran, L., & Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. *Journal of Research on Educational Effectiveness*, *7*, 358–383. https://doi.org/10.1080/19345747.2013.856978

Sauter, M., Uttal, D. H., Alman, A. S., Goldin-Meadow, S., & Levine, S. C. (2012). Learning what children know about space from looking at their hands: The added value of gesture in spatial communication. *Journal of Experimental Child Psychology*, *111*, 587–606.

Scammacca, N., Fall, A. M., Capin, P., Roberts, G., & Swanson, E. (2020). Examining factors affecting reading and math growth and achievement gaps in grades 1–5: A cohort-sequential longitudinal approach. *Journal of Educational Psychology*, *112*, 718–734. https://doi.org/10.1037/edu0000400

Schweig, J. D., Kaufman, J. H., & Opfer, V. D. (2020). Day by day: Investigating variation in elementary mathematics instruction that supports the common core. *Educational Researcher*, *49*(3), 176–187. https://doi.org/10.3102/0013189X20909812

Sengupta-Irving, T., & Vossoughi, S. (2019). Not in their name: Reinterpreting discourses of STEM learning through the subjective experiences of minoritized girls. *Race Ethnicity and Education*, *22*, 479–501. https://doi.org/10.1080/13613324.2019.1592835

Set, A. (2018, February 27). *Study concludes MAP growth items align to Common Core State Standards*. NWEA Blog. https://www.nwea.org/blog/2018/study-concludes-map-growth-items-align-common-core-state-standards/

Shin, T., Davison, M. L., Long, J. D., Chan, C. K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, *23*, 92–100. https://doi.org/10.1016/j.lindif.2012.10.002

Simms, N., & Gentner, D. (2009, April). Helping children find the middle: Spatial language facilitates spatial reasoning. Poster presented at the *Society for Research in Child Development*. SRCD.

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sparks, D. M., & Pole, K. (2019). "Do we teach subjects or students?" Analyzing science and mathematics teacher conversations about issues of equity in the classroom. *School Science and Mathematics*, *119*, 405–416. https://doi.org/10.1111/ssm.12361

Stieff, M., & Uttal, D. (2015). How much can spatial training improve STEM achievement?. *Educational Psychology Review*, *27*, 607–615.

Uttal, D. H. (2000). Seeing the big picture: Map use and the development of spatial cognition. *Developmental Science*, *3*, 247–264.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how?. *Psychology of learning and motivation* (Vol. *57*, pp. 147–181). Academic Press.

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics?. *Current Directions in Psychological Science*, *22*, 367–373.

Visintainer, T. (2020). "I think at first glance people would not expect me to be interested in science": Exploring the racialized science experiences of high school students of color. *Journal of Research in Science Teaching*, *57*, 393–422. https://doi.org/10.1002/tea.21597

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835. https://doi.org/10.1037/a0016127

Yang, W., Liu, H., Chen, N., Xu, P., & Lin, X. (2020). Is early spatial skills training effective? A meta-analysis. *Frontiers in Psychology*, *11*, 1938. https://doi.org/10.3389/fpsyg.2020.01938

Young, C. J., Levine, S. C., & Mix, K. S. (2018). The connection between spatial and mathematical ability across development. *Frontiers in Psychology*, *9*, 755. https://doi.org/10.3389/fpsyg.2018.00755

How to cite this article: Boda, P. A., James, K., Sotelo, J., McGee, S., & Uttal, D. (2022). Racial and gender disparities in elementary mathematics. *School Science and Mathematics*, 122, 36–53. https://doi.org/10.1111/ssm.12506