Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 24-1059

Same Idea, Shifting Standards: An Experimental Study of Racial-Ethnic Biases in Ambitious Math Teaching

Jeannette Garcia Coppersmith
Harvard University

Hannah Kleen
Leibniz Institute for Research and Information in Education

Cynthia Pollard
UnboundEd

Heather C. Hill
Harvard University

Teacher expectations and judgments about student capabilities are predictive of student achievement, yet such judgments may be influenced by salient dimensions of student identity and invite biases. Moreover, ambitious math teaching may also invite teacher biases due to the emphasis on student-generated inputs and ideas. In this pre-registered audit experiment, we investigate teacher biases in a) expectations and judgments about student capabilities in math and b) teacher responsiveness to students' mathematical thinking. Through a between-subjects design, we randomly assigned teachers to a simulated classroom composed of predominantly Black, Latinx/e, or White students and prompted them to respond to a student's mathematical solution. We also prompted teachers to judge the quality of the student's mathematical thinking and rate their expectations about the difficulty of the problem for the typical student. Our findings show teachers expected greater task difficulty in both the Latinx/e and Black classroom conditions relative to the White. We also found teachers may be more likely to support student sense-making and provide more positive, substantive affirmations to Black students relative to White students for the same mathematical solution. We did not find differences by condition in other dimensions. Our findings have implications for teacher training and reform-oriented mathematics instruction.

VERSION: Octoberr 2024

Suggested citation: Coppersmith, Jeannette Garcia, Hannah Kleen, Cynthia Pollard, and Heather C. Hill. (2024). Same Idea, Shifting Standards: An Experimental Study of Racial-Ethnic Biases in Ambitious Math Teaching. (EdWorkingPaper: 24-1059). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/dxy7-eh06

Same Idea, Shifting Standards: An Experimental Study of Racial-Ethnic Biases in Ambitious Math Teaching

Jeannette V. Garcia Coppersmith[a], Hannah Kleen[b], Cynthia Pollard[c], and Heather C. Hill[a]
[a]Harvard Graduate School of Education, Cambridge, Massachusetts, USA; [b]Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany; [c]Stanford University, Stanford, California, USA

Author Note

This paper was funded by the National Science Foundation under grant NSF-1620914. The authors report there are no competing interests to declare. Data available upon request from authors.

ABSTRACT

Teacher expectations and judgments about student capabilities are predictive of student achievement, yet such judgments may be influenced by salient dimensions of student identity and invite biases. Moreover, ambitious math teaching may also invite teacher biases due to the emphasis on student-generated inputs and ideas. In this pre-registered audit experiment, we investigate teacher biases in a) expectations and judgments about student capabilities in math and b) teacher responsiveness to students' mathematical thinking. Through a between-subjects design, we randomly assigned teachers to a simulated classroom composed of predominantly Black, Latinx/e, or White students and prompted them to respond to a student's mathematical solution. We also prompted teachers to judge the quality of the student's mathematical thinking and rate their expectations about the difficulty of the problem for the typical student. Our findings show teachers expected greater task difficulty in both the Latinx/e and Black classroom conditions relative to the White. We also found teachers may be more likely to support student sense-making and provide more positive, substantive affirmations to Black students relative to White students for the same mathematical solution. We did not find differences by condition in other dimensions. Our findings have implications for teacher training and reform-oriented mathematics instruction.

Keywords: Race, equity, mathematics instruction, ambitious math teaching, disparities, teachers and teaching, teacher bias, racial bias, implicit bias, randomized experiment, experimental research

SAME IDEA, SHIFTING STANDARDS: AN EXPERIMENTAL STUDY OF RACIAL-ETHNIC BIASES IN AMBITIOUS MATH TEACHING

"Teachers, like all of us, use the dimensions of class, race, sex, ethnicity to bring order to their perception of the classroom environment" (Lightfoot, 1978, pp. 85-86)

The United States has one of the widest math performance distributions in the world (National Center for Education Statistics, 2019)—a sobering indicator of systemic inequity and social injustice in mathematics education. European American or White[1] and Asian American students consistently demonstrate stronger math achievement compared to Latinx/e[2], Native American, and Black[3] students, contributing to magnified disparities in later educational outcomes, lifetime earnings, and overall access to opportunity in American society (Gutiérrez, 2012; Jencks & Phillips, 2011; Reardon et al., 2015). Scholars have called for a critical examination of "opportunity gaps" for student learning in math, gaps which may help to explain differences in student performance (Flores, 2007; Ladson-Billings, 2006; Martin, 2009; Milner, 2010). Such gaps in opportunity include differences in overall teacher quality, curricula, school funding, as well as educator expectations and mindsets about students of color that limit access to academically rigorous learning experiences (Irvine, 2010; Milner, 2012).

To better understand opportunity gaps, scholars have called for a closer examination of racial-ethnic[4] biases in teachers' instructional decisions (Warikoo et al., 2016). Indeed, disparities between Black and White student achievement in the U.S.—an important measure of education inequality—correlate with teachers' implicit racial biases at the county level, suggesting a relationship between teacher bias, instructional mechanisms, and student achievement (Chin et al., 2020). Findings from audit, experimental, and naturalistic studies have demonstrated racial-ethnic biases favoring dominant student groups, with consistent findings across various instructional dimensions identifying biases in teacher's general academic expectations (Tenenbaum & Ruck, 2007; van den Bergh et al., 2010) and associated student achievement (Peterson et al., 2016), assessments of student ability for ambiguous solutions in math (Copur-Gencturk et al., 2019), grading (Malouff & Thorsteinsson, 2016; Quinn, 2020b), feedback and praise for writing (Harber et al., 2012), general classroom discourse patterns, and disciplinary and educational service referrals (Tenenbaum & Ruck, 2007).

We extend this work in two ways. First, to our knowledge no studies have examined how teachers' expectations of task difficulty–teachers' prospective judgments about how hard a task will be for students–might differ when they are cued to think about students of different race-ethnicities. Critically, expectations about future student learning not only predict student achievement (Weinstein, 2002), they also vary between students of different race/ethnicities, with lower expectations for students belonging to historically minoritized groups (Peterson et al., 2016; Pit-ten Cate & Glock, 2023; Rubie-Davies et al., 2006). We argue that assigning tasks to students is a concrete mechanism by which expectations can introduce bias into students' opportunities to learn and investigate those difficulty expectations here. Indeed, prior research has found racial bias against Black students in teacher task selection (Pollard, 2022), making difficulty estimations a fruitful line of study.

Second, few studies have investigated whether biases arise in teachers' in-the-moment decisions about instruction in the domain of mathematics, despite teachers having wide discretion about such decisions (Ball, 2018). We argue that teacher biases may be particularly relevant to *ambitious instruction*, a set of instructional practices that relies heavily on dialogue between teachers and students to support

---

[1] In line with prior scholarship by Ewing (2020) and others, we capitalize "White" in order to underscore "the specificity and significance of Whiteness–the things that it is, the things that it does."

[2] We use the gender-neutral and inclusive pan-ethnic identifier "Latinx/e" while understanding that not all members of this group may self-identify with this term (Trejos, 2024).

[3] We use the term "Black" in order to recognize this racial-ethnic identity in a transnational manner, inclusive of individuals with roots in the Caribbean and Latin America (Laws, 2020).

[4] Our use of the term "racial-ethnic" aligns with prior scholarship on the interrelationship between racial and ethnic identities. We take race to define a "socially constructed lay theory" (p. 4) representing perceived phenotypic differences between groups of people. Similarly, ethnicity represents national origin and the cultural traditions of social groups. Race and ethnicity are socially constructed and highly correlated both in terms of perception and in terms of sociocultural identity. Further, while terms like "Latinx" represent an ethnic social categorization, terms like "Black" often represent both racial and ethnic social categories. As such, we take the term "racial-ethnic" to capture the overlapping meanings of these constructs (Cross & Cross, 2008).

student engagement in cognitively demanding work (Lampert et al., 2013). Ambitious instruction includes teacher instructional moves such as eliciting student thinking, using that thinking in subsequent instruction, and affirming the substance of students' mathematical ideas—what some call responsive teaching in mathematics (Kang, 2022; Richards & Robertson et al., 2015). In such classrooms, student sense-making about mathematics is "put into play" as a resource for learning, and teachers respond by affirming and building on those ideas. Such centering of student sensemaking as a resource for learning is associated with improved learning outcomes in math, as it supports higher-order thinking and the construction of conceptually-based subject matter knowledge (Carpenter et al., 1989; Cohen & Hill, 2000; Schoenfeld et al., 2023; Smylie & Wenzel, 2006). However, as Ball (2018) notes, the high volume of teacher decisions during ambitious math instruction—and the wide latitude for discretion—may create more opportunities for teacher biases to emerge.

We investigate the relationship between student race and teacher expectations and responsiveness via a three-factorial, between-subjects design in which we randomly assigned teachers to a simulated classroom environment composed of predominantly Black, Latinx/e, or White students and then prompted them to verbally respond to a hypothetical target student's mathematical solution. We also prompted teachers to rate the quality of the hypothetical student's mathematical thinking and the level of difficulty of the problem for the typical 4th grader in the U.S. Our specific research questions are: What is the extent to which teachers demonstrate biases amongst Black, Latinx/e and White students in their…

1) … expectations and judgments about student math capabilities, including a) teacher judgments about the quality of student mathematical thinking and b) teacher expectations of task difficulty for typical students?
2) … instructional moves that are responsive to students' mathematical solutions, including elicitation and use of student thinking, substantive affirmations to students, and student-centered instructional moves?

We argue that this study combines the clear causality that arises from audit experiments with more authentic depictions of classroom life and teacher decision-making (Herbst et al., 2011; Charalambous, 2020). Because ambitious math instruction and high expectations are linked to improved learning outcomes for students, knowledge about the extent to which student racial-ethnic groups may experience such opportunities to learn differentially in these high-leverage instructional domains will contribute to the field's understanding of how racial achievement disparities come to be at the classroom level. More importantly, a concrete understanding of the instructional mechanisms behind racial inequality has the potential to inform teacher education and professional development with specific areas to intervene on, for example, through teacher coaching and rehearsals (Mancenido et al., 2025)

BACKGROUND

Here, we discuss the prior literature on racial-ethnic biases in teacher expectations and instruction through the lens of implicit social cognition, with particular attention to domain-general biases and those specific to the mathematics instructional context. We follow with an overview of ambitious mathematics instruction and its relationship to student learning, arguing that this type of teaching may simultaneously instigate and inhibit student opportunities to learn according to how these dimensions of instruction interact with teacher biases. We articulate the gap we aim to address before arriving at the present study.

RACIAL-ETHNIC BIASES IN THE CLASSROOM

Observational studies suggest minoritized learners miss out on grade-level appropriate assignments, high-quality instruction, deep engagement, and high expectations for their learning (TNTP, 2018). Classrooms serving students of color often exhibit reduced cognitive challenge, with a greater emphasis on procedural skill or rote practice, particularly for Black and Latinx/e students in urban settings (Boston & Wilhelm, 2017). An examination of racial-ethnic biases in the classroom through the lens of social psychology may help to explain such disparities in student opportunities to learn (Warikoo et al., 2016).

Specifically, implicit social cognition theory, which refers to the subconscious influence of one's past experiences on their perceptions, attitudes, and behaviors (Greenwald & Banaji, 1995), provides a conceptual framework for understanding the ways in which racial-ethnic biases may shape teachers' expectations, judgments about student capabilities, and responsiveness to student solutions. Biases in behavior and judgments trace back to stereotypes or cultural beliefs about the behavior and traits of social groups (Eagly & Chaiken, 1993; Hilton & von Hippel, 1996; Tajfel, 1970). A function of stereotypes is group categorization, which is used as a way to save time and cognitive resources when processing information and making sense of one's complex social world (Schneider, 2004).

In education, dominant social narratives about Black and Latinx/e students have the potential to inform stereotyped perceptions of minoritized learners. Racial-ethnic minority groups have been harmfully mischaracterized as having low intelligence or being less hard working (Bonefeld et al., 2020; Fiske et al., 2007). Stereotypes about minoritized learners may indeed shape individual's perceptions; in a randomized experiment, Quinn (2020a) found that viewers who watched a TV news story about racial achievement gaps expressed more exaggerated stereotypes of Black Americans as lacking education and showed increased implicit stereotyping of Black students as less competent compared to their White counterparts. Common stereotypes in education concerning minoritized learners, particularly in math, are enduring and have the potential to shape teacher perspectives of their students and decisions in the classroom (Martin, 2009).

Paradoxically, stereotypes that lead teachers to hold lower expectations may result in more positive judgments about specific individuals' work. This occurs, as described by the shifting standards theory, when stereotypes about social groups inform standards for evaluating members of that group, and those standards shift and differ across groups (Biernat et al. 1991; Biernat, 1995). Marshall's (2020) discussion of the use of terms like "well-spoken" or "articulate" in reference to Black people simply communicating effectively helps to illustrate the shifting standards theory. Such descriptors imply a low standard for members of minoritized groups and result in more positive judgments about the same skill or ability as compared to those for white individuals.

BIASES IN TEACHERS' ACADEMIC EXPECTATIONS AND JUDGMENTS ABOUT STUDENT ABILITIES

Teacher expectations are the cognitive, "inferential judgments that teachers make about probable future achievement and behavior of a student" (Brophy & Good, 1974, p. 129). In mathematics, expectations and judgments about student abilities can involve teacher expectations of the difficulty of a problem or task for a student and judgments about the quality and sophistication of students' mathematical thinking. Informed by implicit stereotyped perceptions, teacher expectations differ across student groups, as they are influenced by characteristics such as the student's socioeconomic status, gender, (dis)ability, language, and race-ethnicity—the focus of this study (Rubie-Davies et al., 2012). Prior research provides evidence on this point: a meta-analysis of 32 studies concluded teacher expectations are higher for White students compared to Black ($d=.25$) and Latinx/e ($d=.46$) students (Tenenbaum & Ruck, 2007). Non-Black teachers have also been found to hold significantly lower expectations of Black students compared to Black teachers (Gershenson et al., 2016). Further, an experimental audit study by Anderson-Clark et al. (2008) found that teacher ratings of students' motivation and achievement-related behaviors were significantly lower for a Black-associated name compared to a White-associated name, but no main effect was found from explicitly-stated descriptions of student race. These studies highlight the unique role of implicit racial biases over explicit ones in influencing teacher expectations and judgments of student academic abilities.

Such differential expectations by race appear consequential for student achievement. Teachers' implicit biases have been shown to predict racial achievement gaps between Dutch-origin and Turkish immigrant students, and this relationship is mediated by teacher expectations (van den Bergh et al., 2010). Peterson and colleagues (2016) found implicit biases in favor of ethnic-majority student groups in New Zealand likewise predicted student achievement in reading and math, and that students in classrooms of teachers who explicitly held higher expectations for their learning showed greater improvements in reading. While some scholars have argued that differential teacher expectations by student race merely reflect racialized achievement disparities (Jussim & Harber, 2005), Peterson and colleagues' models (2016) identified the effect of bias while also controlling for prior student achievement.

In mathematics, there is mixed evidence for biased teacher expectations for and judgments about student abilities. Using a fractions task involving adding unlike fractions, Pollard (2022) found teachers were more likely to attribute errors to the student, rather than to the student's prior instruction, for Black students as compared to White students, and this effect was larger when the teacher held stronger pro-White implicit bias (as measured by the Teachers' Implicit Association of Academic Achievement Task (TIAAAT) adapted from Peterson et al. (2016)). Examining interactions between student gender and race-ethnicity, Copur-Gencturk and colleagues (2019) found that while teachers were unbiased when rating the correctness of student solutions, they demonstrated racial-ethnic biases in their ratings of student math abilities for partially correct solutions. Teacher ratings were lowest for Black and Latinx/e or Hispanic girls compared to boys and White students, underscoring the additive effects of marginalized identities for inviting biases. However, no evidence of racial bias was found on average in a study that assigned race-specific names to student math solutions (Copur-Gencturk et al., 2023). In a similar experiment carried out in the virtual context, this time using screenshots of students explaining their work, teachers did not demonstrate correctness or ability bias, but they were more likely to recommend male students for a gifted program and refer Black students for special education based on identical student work (Copur-Gencturk et al., 2022). These findings demonstrate the role of biases in teachers' subjective judgments about students' mathematical thinking, particularly at the intersections of marginalized identities of student race and gender.

BIASES IN TEACHERS' INSTRUCTION

Biases have also been shown in various dimensions of teachers' instruction. In their meta-analysis of both experimental and non-experimental studies on teacher speech patterns, Tenenbaum and Ruck (2007) found racial biases in teachers' valence of speech and referrals, with more positive speech (e.g. encouragement) and more neutral speech (e.g. questions) for European American students as compared to Black and Latinx/e students. Teachers were found to direct negative speech equally across student groups. While the associations were small, they were statistically significant within the meta-analytic framework, in line with previous research on the small-to-moderate effects of implicit biases. However moderate the effect sizes may be, the consequences of racial-ethnic biases have the potential to be cumulative and harmful (Greenwald et al., 2014) over the course of a student's entire K-12 schooling history, leaving them underprepared for higher education.

Scholarship has also uncovered racial biases in teachers' subjective feedback on student writing. In an experimental audit study of student essays, Harber et al. (2012) found that White teachers demonstrated positive feedback bias toward Black and Latinx/e students, providing less critical feedback than they did for White students, but only on the more subjective components of writing like content rather than mechanics. In line with the shifting standards theory described above, the authors of this study hypothesize that teachers may have provided more positive affirmations on Black and Latinx/e students' writing because they had lower initial expectations for their writing skills. While prior research shows teachers provide biased ability ratings towards Black and Hispanic students' work (Copur-Gencturk et al., 2019), there are no studies to date, to our knowledge, that examine biases in teachers' affirmations of student thinking in mathematics.

Research on racial biases in math teaching has begun to emerge. In the area of mathematical task selection—which reflects teacher judgments about student abilities, a focus of this study—Pollard (2022) found that teachers were more likely to select lower-quality math curricular materials and remedial instruction for Black students compared to White students. Pollard (2022) also found that teachers were more likely to endorse student-centered instructional moves for Black students compared to White students. Finally, student race showed no relation to the grades or feedback provided in mathematics.

Such findings align with prior results from audit experiments in psychology in which biases have been found to inform behavior and judgments when respondents are given greater latitude for interpretation. Indeed, in the absence of information and in ambiguous situations that require subjectivity, student group membership may become more salient, causing implicit biases to play a more pronounced role in informing action (Dovidio, 2001). In ambitious mathematics teaching—where teachers must do more than simply evaluate student answers as correct or incorrect and instead interpret student thinking and articulate a reasoned reaction—implicit biases may be more likely to affect teachers' cognitive processing. Yet we know of no studies to date that examine racialized differences in teacher responsiveness in the moment in this crucial instructional domain.

AMBITIOUS MATHEMATICS INSTRUCTION AND TEACHER RESPONSIVENESS

Greater cognitive activation on the part of the student is core to what some term "ambitious math instruction" (Lampert et al., 2013; Smylie & Wenzel, 2006). In line with reforms outlined by National Council of Teachers of Mathematics (NCTM, 2000, 2014) and Common Core State Standards Initiative (2010), this approach to instruction emphasizes the rigor of mathematical tasks and student sense-making around those tasks. Teacher responsiveness to student thinking can contribute to student sense-making, making responsive teaching—sometimes called dialogic teaching (Alexander, 2008)—in mathematics an important tool for building student knowledge. In this form of instruction, teachers work to surface students' mathematical ideas, have students justify their thinking processes, and have them attend to each other's reasoning. Such centering of student sensemaking as a resource for learning is associated with improved learning outcomes in math, as it supports higher-order thinking and the construction of conceptually-based subject matter knowledge (Carpenter et al., 1989; Schoenfeld et al., 2023; Smylie & Wenzel, 2006).

Effective responsive teaching in mathematics necessitates a variety of instructional moves from the teacher as they work with student mathematical thinking in the moment (Boston, 2012; Stein et al., 2008). Key responsive teaching moves investigated here include eliciting, using, and affirming student thinking as well as supporting student sensemaking about the math content:

- *Elicitation* refers to the practice of surfacing specific students' mathematical ideas, and can involve prompting a student to explain their strategies or other ways of "getting student thinking on the table" (Shaughnessy & Boerst, 2018).
- *Use of student thinking* in mathematics refers to the teacher keeping a student's idea in play as an object of discourse and includes the teacher or another student revoicing, emphasizing, restating, and probing a student's idea or solution process (Herbel-Eisenmann et al., 2013; Jacobs & Empson, 2016; Lampert, 2001; O'Connor & Michaels, 1996).
- *Affirming* student thinking includes teacher praise and "affirmation of something a student did well" (Brophy, 1981; Sun & Ruef, 2023).
- Finally, *supporting student sensemaking* refers to instructional moves in mathematics that create opportunities for students to have ownership over and make sense of the content, for example, by engaging students in a think-pair-share.

As evidenced by the variety of instructional moves involved in responding to and assessing student thinking in-the-moment, ambitious math teaching is inherently complex. It opens the door to the impossibly vast realm of student ideas and ways of thinking about mathematics, introducing greater possibilities for instructional content and consequently a greater degree of uncertainty. Greater attention to student thinking is necessitated, as teachers must attend to both the content standards and student-generated inputs; teachers must be accountable to students' sometimes unconventional intellectual contributions as well as to the established standards of the discipline (Lampert, 1990). On one hand, surfacing the array of potential student ideas in the math classroom can create new opportunities to reveal student thinking that would otherwise go unacknowledged, making learning experiences more equitable. On the other hand, contributions from minoritized learners may be undervalued in the interactional space of the math classroom, thereby reproducing racial hierarchies (Gutiérrez, 2013; Martin, 2009; Nasir et al., 2008).

THE PRESENT STUDY

Findings from the literature on racial-ethnic biases both in general and in the domain of mathematics have consistently demonstrated lower expectations and perceptions of student ability for minoritized students of color, work we replicate here and extend in four important ways. First, we aim to replicate prior findings regarding teacher expectations and judgments about student ability. Second, we examine how teachers' expectations of task difficulty—their anticipation about how hard a task will be for students—might differ when they are cued to think about students of different races or ethnicities. Prior work has shown biases in teachers' curricular selection in the mathematics classroom (Pollard, 2022), and because such task selection is informed by expectations for subsequent student learning, we extend this work by investigating biases in teacher expectations of task difficulty for the student in the context of discursive instructional space. Third, we extend prior studies about biases in instruction to the domain of responsive math teaching, a domain that requires significant teacher decision-making and discretion and thus opens the potential for bias (Ball, 2018). Fourth, because we focus on ambitious instruction, we also feature non-standard, innovative student solution methods, thus extending existing studies of teacher bias beyond correct and partially correct student work. We do this work using a pre-registered experimental design, strengthening the study's internal validity, and via scenarios representing typical K-8 mathematics classrooms that elicit teachers' verbal, in-the-moment responses, which contributes an authentic depiction of classroom life and teacher decision-making (Herbst et al, 2011; Charalambous, 2020).

METHODS

We measured teacher judgments about the quality of students' mathematical thinking, teacher expectation of task difficulty, and teacher responsiveness to student solutions in the context of six animated teaching simulations (Herbst et al., 2011; Charalambous, 2020), here called vignettes, and operationalized student race-ethnicity using visual cues and names to understand how teachers' responses to and thinking about student work in the vignettes might systematically vary. We describe our methods in more detail here.

SAMPLE

We recruited participants using a Qualtrics teacher panel in the U.S. from June through August of 2022. Our screening criteria selected for K-8 teachers who had taught mathematics within the last 3 years. Qualtrics returned a data set of 315 teachers, from which we omitted three teachers who failed to authentically engage in the survey (e.g. had blank, incomplete, or irrelevant responses). There were 19 teachers who indicated they currently taught high school that we retained in our sample. Our survey included attention checks throughout, with over 99% of teachers passing at least one. Because an inspection of responses revealed that individuals who failed an attention check did offer appropriate and relevant verbal responses to the vignettes, we did not remove their responses. Our final sample of *N*=312 teachers is shown in Table 1. In line with the teacher population in the U.S., the majority of our sample identified as White women. Our sample, however, is weighted toward more experienced educators than is typical in the U.S. (National Center for Education Statistics, 2023).

Table 1. Sample Characteristics for Full Sample and Across Conditions

	Full Sample *N*=312	Black Condition *n*=109	White Condition *n*=106	Latinx/e Condition *n*=97	Pearson Chi-square and ANOVA
	N (%)	*n(%)*	*n(%)*	*n(%)*	
Gender					
Woman	272 (87)	95 (87)	88 (83)	89 (92)	$\chi^2(2) = 3.46$, p = 0.18
Man	39 (13)	14 (13)	18 (17)	7 (7)	$\chi^2(2) = 4.43$, p = 0.11
Nonbinary	1 (<1)	0 (0)	0 (0)	1 (1)	$\chi^2(2) = 2.22$, p = 0.31
Teaching experience					$\chi^2(12) = 10.5$, p = 0.57
<1 year	3 (1)	1 (1)	2 (2)	0 (0)	
1-2 years	18 (6)	7 (6)	5 (5)	6 (6)	
3-5 years	34 (11)	14 (13)	6 (6)	14 (14)	
6-10 years	66 (21)	25 (23)	22 (21)	19 (20)	
11-15 years	46 (15)	18 (17)	18 (17)	10 (10)	
16-20 years	41 (13)	13 (12)	13 (12)	15 (15)	
20+ years	105 (34)	31 (28)	40 (38)	33 (34)	
Lowest Grade taught					$\chi^2(6) = 4.1$, p = 0.66
PreK-2	82 (26)	24 (22)	34 (32)	24 (25)	
3-5	139 (45)	51 (47)	45 (42)	43 (44)	
6-8	72 (23)	28 (26)	22 (21)	22 (23)	
9-12	19 (6)	6 (6)	5 (5)	8 (8)	
Teacher race-ethnicity					
Black	22 (7)	3 (3)	14 (13)	5 (5)	$\chi^2(2) = 9.74$, p = 0.01
Latinx/e	16 (5)	7 (6)	4 (4)	5 (5)	$\chi^2(2) = 0.78$, p = 0.69
Native Am.	3 (1)	0 (0)	1 (1)	2 (2)	$\chi^2(2) = 2.29$, p = 0.21
AAPI	11 (4)	2 (2)	5 (5)	4 (4)	$\chi^2(2) = 1.46$, p = 0.51
White	239 (77)	88 (81)	73 (69)	78 (80)	$\chi^2(2) = 5.36$, p = 0.07
Multiracial	21 (7)	9 (8)	9 (8)	3 (3)	$\chi^2(2) = 2.97$, p = 0.20
School Demographics	mean				
Urbanicity	0.29	0.31	0.30	0.27	$\chi^2(2) = 2.43$, p = 0.30
Percent White	0.55	0.55	0.53	0.57	$\chi^2(2) = 2.66$, p = 0.26
SES average	0.28	0.32	0.20	0.34	$\chi^2(2) = 4.83$, p = 0.09

Avg. Math Ach	0.03	0.02	-0.005	0.06	$\chi^2(2) = 3.21$, p = 0.20

Note: Column percentages are shown in parentheses. Means are provided for school demographic variables. A Fisher's exact test was used for categorical variables with few frequencies in cells, and Bartlett's equal-variances test was used for continuous variables.

STUDY DESIGN

We employed a three-factorial, between-subjects design in which participants were randomly assigned to view vignettes from a hypothetical elementary classroom composed of predominantly Black, Latinx/e, or White students. We composed vignettes of between 5 to 11 slides or frames, and were identical save for the race-ethnicity of the students. Each vignette showed study participants an elementary math problem introduced by a 'teacher;' student work time on the problem; and a "target" student solution to the problem which provided either their answer, their thinking, or both. All target student responses could be interpreted as correct, several featured relatively sophisticated student mathematical reasoning, but as the relevant target student utterances in Table 2 show, all responses invited further probing or discussion of the student's thinking. The final frame of the vignette asked study participants to verbalize what they would say to the class next into their device's microphone, and then to verbalize the instructional step they would take after that initial comment.

Table 2. Item Descriptions and Target Student Responses in the Six Vignettes

Item	Mathematical problem or prompt	Instructional scenario	Relevant target student utterances
1	Represent 2/3 on a number line	Students work independently. Teacher brings class together and asks two students to share their answers.	*S1*: First I divided the part of the number line between 0 and 1 into 3 equal segments. Then, I shaded two of them. So ⅔ is right here. *S2*: I did the same thing. I just shaded in two different segments. (Shows shading starting at ⅓ and ending at 1).
2	T has 14 papers. J has 9 papers. How many fewer papers does J have than T?	Students work independently. Teacher asks a student to share thinking. Student replies using "counting up" method.	*S*: I used my fingers and counted up from 9 to 14. That's 10, 11, 12, 13, 14. That's 5 counts, so J has 5 papers fewer than T.
3	What fraction of the rectangle is shaded? (Rectangle shows ⅖ shaded)	Students think independently and then turn and talk about problem. Teacher elicits a response from a student. Student replies with a numeric answer and a solution.	*S*: Two fifths. There are five parts and two of them are shaded.
4	12 x 5 =	Teacher elicits student response. Student replies with a non-standard method.	*S*: I did it in my head. First I multiplied 5 times 10 and got 50. Then I multiplied 5 times 2 and got 10. Then I added 50 and 10 and got 60.
5	19 + 6 =	Teacher asks for a student to come to the board to solve problem. Student volunteers and solves the problem using a non-standard method.	*S*: I took away a one from the 6 and added it to the 19. Then it was easy, 5 plus 20 is 25.
6	How many tens does 743 have?	Teacher first asks how many ones 743 has. Student answers 3, and	*S1*: I think it has four tens.

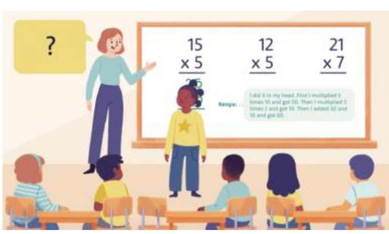
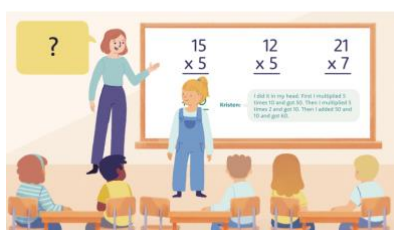
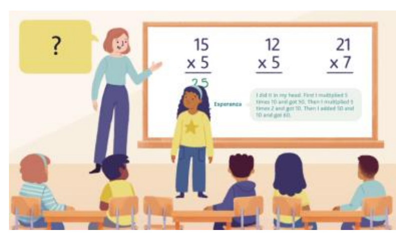
	teacher affirms answer as correct. Teacher then asks how many tens 743 has. Teacher calls on two students.	*S2*: I think it has seventy-four tens.

To operationalize race-ethnicity, four of the six students present in each vignette signaled either Latinx/e, Black, or White students in terms of physical attributes like skin tone and hair features. Further, the student speaking had a name commonly associated with that particular group to further signal their racial-ethnic identity. As is typical for U.S. classrooms, the teacher in all vignettes was White (National Center for Education Statistics, 2023). See Figure 1 for the last frame of a vignette across the three conditions.

Teachers were told that the purpose of the study was to pilot the vignettes as an instrument to capture teachers' responses to classroom situations (see Authors, 2024, under review). Randomly assigning teachers to experience vignettes in only one condition (a between- rather than within- subject design) diminished the potential for social desirability bias (Hofmann et al., 2005). Overall, 87% of teachers correctly noted the race of the student speaking in our manipulation checks—86% in the Black condition, 82% in the Latinx/e condition, and 92% in the White condition. We found that emergent evidence to suggest teachers in the White classroom condition may have been more likely to correctly notice and report the race of students in the vignettes (χ^2=4.65, *p*=0.10).

We acknowledge that student identities are complex and intersectional, and that race and ethnicity cannot be fully disentangled. While the categories we use are certainly not necessarily mutually exclusive in the real world, e.g. in the case of Afro-Latinx students, our purpose was to attempt to make the student's racial-ethnic identity salient in order to measure teacher biases. We accept this trade-off between the study's simplification of student race-ethnicity and the benefit of an experimental design.

Figure 1*Three experimental conditions as seen in last frame of vignette 4*

Black: *Martin Luther King Jr. Elementary School*	White: *Eugene O'Neill Elementary School*	Latinx/e: *Sonya Sotomayor Elementary School*
		

Note: The teacher utterance preceding the student response in this hypothetical teaching scenario was: *"Let's look at the next one. Who can solve 12 x 5?"*

PROCEDURE

Vignettes were presented to teachers via the survey software Qualtrics. We fixed the first vignette to randomize teachers into their condition, and the subsequent five were presented in random order to avoid potential order effects. We note that by chance, Black teachers were over-represented in the White condition (χ^2(2) = 9.74, *p* = 0.01), and White teachers were slightly over-represented in the Black condition (χ^2(2) = 5.36, *p* = 0.07). After respondents viewed the frames depicting the sequence of instruction in each vignette, the final frame demonstrated the target student response(s) and directed participants to "Speak what you would say next to the class" as if they were the hypothetical teacher (Question 1, Q1). Next, they were asked a second question: "After saying this to the class, what would you do next?" (Question 2, Q2). We captured teacher responses using an embedded voice transcription software, *Phonic*.

After recording their responses to the open-ended Q1 and Q2 questions, teachers responded to closed-ended Likert-type items pertaining to each vignette using a 1 (low) - 5 (high) scale. The survey concluded with questions limited to teacher background. In addition to attention checks, we included manipulation checks throughout the survey to determine whether respondents noticed the race/ethnicity of the students in the vignettes.

MEASURES

See Table 3 for a summary of the prompts and nature of data for all pre-registered measures. We include, for completeness, our measure of eliciting student thinking, which we were unable to analyze because of poor fit between the items and construct.

Table 3. Constructs Measured and Prompts in the Survey

Construct	*Survey Prompts*
Teacher Judgments About the Quality of Students' Mathematical Thinking (Likert, 1-5 scale)	How mathematically sophisticated is [student]'s response? How likely is it that [student] has a solid conceptual understanding of [problem type]? [Student]'s response indicates they are on track toward mastering [problem type]. [Student]'s thinking is mathematically sound. The teacher should encourage [student] to solve similar problems in the same way. The teacher should encourage the class to solve similar problems in the same way.
Task Difficulty (1-5 scale)	How difficult would this problem be for the typical 4th grader in the U.S.?
Eliciting Student Thinking	Q1: Speak what you would say next to the class.
Using Student Thinking	Q1: Speak what you would say next to the class.
Teacher Affirmation	Q1: Speak what you would say next to the class.
Supporting Student Sense-Making	Q2: After saying this to the class, what would you do next?

Teacher Expectations and Judgments About Students' Math Abilities

Teacher Judgments about the Quality of Student Mathematical Thinking. We measured teacher judgments about the quality of the student's mathematical thinking via six Likert-type items (1-5 scale) per vignette where the teacher rated the responses' sophistication, the student's likely level of understanding, their content mastery, the soundness of their thinking, and the extent to which the teacher would encourage the student and class to use the approach described by the target student (an endorsement of the student's strategy). This measure had acceptable internal reliability (α=0.90) across the 36 items. We constructed and used scores output by a nested measurement model, with the six ratings under each vignette, and a confirmatory factor analysis revealed adequate fit ($\chi^2(580) = 1224.49$, $p<0.00$ CFI = 0.93 TLI = 0.92 RMSEA = 0.06 SRMR = 0.07; Hu & Bentler, 1999).

Teacher Expectations of Task Difficulty for the Student. We measured teacher expectations for task difficulty with one item per vignette intended to capture how difficult they expected the math problem to be for "the typical fourth grader." Teachers responded on a 5-point Likert scale, with 1 meaning not difficult at all and 5 being very difficult. Teachers' reported ratings of difficulty were cohesive across the six vignettes (α= 0.77), and a confirmatory factor analysis revealed adequate fit ($\chi^2(6) = 7.32$, $p = 0.29$, CFI=1.00, TLI=1.00, RMSEA = 0.03, SRMR = 0.03; Hu & Bentler, 1999).

Teacher Responsiveness to Student Thinking

Before analysis, all open-ended responses were transcribed and reviewed against their original audio by human raters; the ASR transcriptions were 95% accurate at the word level and the meaning of most inaccurate words could be inferred. Members of the project team developed an initial set of low-inference codes for the teacher responsiveness measures and completed preliminary coding on pilot data to achieve 80% inter-rater agreement or better, refining these codes to improve clarity. In rare cases, we coded responses as N/A when there was no teacher utterance or it was not responsive to the prompt. Once codes were finalized, the lead team recruited and trained raters using master scores from the pilot data. Raters specialized in a particular code and double-scored responses within each vignette until an average 80% inter-rater agreement rate or better was achieved. The first author held weekly norming sessions for the rating team to discuss and resolve discrepancies. Raters went on to single score responses, and a member of the project team double-scored 20% of those scores to assess rater drift and accuracy. The project team resolved any disagreements that occurred during this process. The Appendix provides codes for all teacher responsiveness measures.

Eliciting Student Thinking (EST). We coded teacher responses to Q1 for the extent to which teachers elicited target student ideas, assigning a zero for each vignette when there was no elicitation, a one for minimal, and a two for strong elicitation of student thinking. The inter-rater agreement rate was 0.90 for EST; however, while a factor analysis indicated a distinct factor for these items (see Table 4), the six vignettes together exhibited poor score reliability, with α=0.34 over the six items. We thus excluded this construct from further analysis.

Using Student Thinking (UST). We coded teacher responses to Q1 for the extent to which teachers used student thinking in their subsequent utterance, assigning a code of zero when the teacher responded in a pro forma way (e.g. “Great work”); a one when they used some aspect of student thinking, such as brief restatement; and a two when there was strong use of student thinking, such as the teacher weaving the student idea into the development of the mathematics. The interrater agreement for UST was 0.78, and we estimated α=0.73 over the six items, demonstrating adequate internal consistency.

Supporting Student Sensemaking (SSS). We coded teacher responses to Q2 for supporting student sense-making, assigning a code of zero when there was no support for student sensemaking in the next planned instructional step (e.g. “I would move on to a new, similar problem”), and a one when the teacher centered student sense-making (e.g. “I would ask who agrees with the answer and to explain why or why not”). The interrater agreement rate for SSS was 0.84, and we estimated α=0.73 over the six items, demonstrating adequate internal consistency.

After scoring for the three constructs above, we conducted exploratory factor analysis that permitted these scores to load onto all three hypothesized latent factors in an unrestricted measurement model. This analysis revealed the strongest standardized factor loadings generally corresponded with their respective hypothesized factors. We fit an adequate unrestrained 3-factor model, $\chi^2(102) = 142.21^{**}$; RMSEA = 0.04; CFI = 0.95; TLI = 0.93; SRMR = 0.12; WRMR = 0.75, which suggests a distinct factor solution for each respective construct.

Teacher Affirmation (TA). We coded Q1 responses for teacher affirmation, assigning a code of zero when there was no affirmation (e.g. “Who agrees?”), a one for a general affirmation (e.g. “Excellent job on that problem”), and a two for specific affirmation (e.g. “Great work explaining there are ten tens hidden in each hundred”). Rater agreement was 0.95 for teacher affirmation (TA). We conducted a separate factor analysis for this affective construct. For the six vignette scores for teacher affirmation in Q1 (responding to the target student directly), we fit an adequate measurement model ($\chi^2(9) = 16.00$, p=0.07, RMSEA=0.05, CFI= 0.99, TLI=0.98, SRMR=0.04, WRMR=0.59). Teachers’ coded responses for TA were also cohesive (α= 0.77). These findings support a distinct factor solution for teacher affirmation.

For all coded measures of teacher responsiveness that were adequately internally consistent—UST, SSS, and TA— we estimated scores for each teacher using 2-parameter logistic (2PL) models using Item Response Theory’s (IRT) Graded Response Model (GRM) for polytomous data (Samejima, 1969).

Other Measures

Teacher background. We use an ordinal variable to measure teaching experience: less than 1 year (0), 1-2 years (1), 3-5 years (2), 6-10 years (3), 11-15 years (4), 16-20 years (5), and >20 years (6). For grade level, we also use an ordinal variable: Pre-K to grade 2 (0), mid-elementary or grades 3-5 (1), middle school or grades 6-8 (2), and high school or grades 9-12 (3). For respondents who indicated multiple grade levels, we used their lowest level taught to obtain a conservative proxy for level of teacher mathematical curricular knowledge. Teacher gender was measured as woman, man, or non-binary. Finally, we used dummy variables to represent the following self-reported teacher identities: Black, Latinx/e, White, Asian American or Pacific Islander, Native American, or multiracial. Due to small sample sizes for some of these identities, we cannot conduct interaction analyses between condition and teacher race; these small sample sizes and a lack of theory also prevent us from interpreting teacher race in the results below.

Teaching context covariates. Some have argued that racial biases in teacher beliefs and behaviors may simply reflect their experiences with racial disparities in student achievement (Jussim & Harber, 2005; Jussim et al., 1996). Teacher expectations are informed by prior student achievement, socioeconomic status, gender, student diagnostic labeling, and student race-ethnicity (Peterson et al., 2016). As such, we asked teachers for their school zip code and cross-referenced that with the Stanford Education Data Archive (SEDA) to obtain the following covariates: school urbanicity as measured by the proportion of students in city or urban locale schools, the percentage of White students in the locale, a composite score describing the metro locale's socioeconomic status, and finally a cohort scale mean of the math achievement in the metro area (Fahle et al., 2021).

ANALYSIS

Due to the experimental nature of the study, we preregistered on OSF. We investigated the potential for bias in teachers' scores on the above dimensions by fitting a series of multiple regression models with student race-ethnicity entered as a fixed effect, with the reference group being the White classroom condition. We used structural equation modeling (SEM) with maximum likelihood and diagonally weighted least squares estimation to model the impact of condition on teacher judgments about student mathematical thinking and expectation of math task difficulty, as this technique allowed us to account for measurement error in the construction of our latent variables (Kline, 2016). We modeled the impact of condition on our responsiveness measures by using multiple regressions, improving the precision of our estimates by including the demographic and covariates variables described earlier in both the SEM and multiple regression models (Equation 1).

Equation 1 $\quad Y_j = B_{0j} + B_1 \mathit{Black}_j + B_2 \mathit{Latinx}_j + B_3 X_j + \mathrm{error}_j$

Where j=participant, X represents the set of covariates described above, and error describes a random error variance term. B_1 and B_2 are the coefficients measuring the main effects of bias. A retrospective power analysis revealed we were between a power estimate of 0.17-0.49 to find the effects we found, which we take as substantively meaningful differences in teacher responses to students. Descriptive analyses and regressions were completed using Stata version 17.0, and factor analysis was completed using RStudio version 2023.12.1+402 with the lavaan package for robust model fit indicators to accommodate polytomous response data.

RESULTS

TEACHER EXPECTATIONS AND JUDGMENTS ABOUT STUDENT MATH ABILITIES

We did not find significant differences by condition in teacher judgments about the quality of students' mathematical thinking (see Table 4). For teacher expectations, our findings show respondents expected greater task difficulty for the typical U.S. student on average when assigned to the Black (β = 0.11, p<.05) and Latinx/e (β = 0.14, p<.05) conditions, $\chi^2(76)$= 64.47, p=0.82, CFI=1.00, TLI=1.00, RMSEA=0.00, SRMR = 0.020. Our findings also showed that teachers working with higher grade levels on average expected greater task difficulty for students (β = 0.14, p<.001). We also found that women (β = -0.12, p<.10), those with more teaching experience (β = -0.05, p<.01), and those teaching in contexts with higher aggregate math achievement (β = 0.20, p<.10) expected lower task difficulty for students, on average.

TEACHER RESPONSIVENESS TO STUDENT MATHEMATICAL EXPLANATIONS

We did not observe significant differences by condition in teacher use of student thinking. Those with more years of teaching experience were more likely to use student thinking (β = 0.07, p<.05), and Black (β = -0.36, p<.10) and Native American educators (β = -1.37, p<.01) were somewhat less likely to use student thinking, but we interpret these coefficients with caution as our sample was insufficiently powered for analyses of teacher racial-ethnic subgroups. When articulating their next instructional steps, we find marginally significant evidence to suggest teachers were more likely to support student sensemaking for Black students relative to White students (β = 0.20, p<.10). Women were also more likely to support student sensemaking compared to men and nonbinary educators (β = 0.31, p<.05). We also find emergent evidence suggesting teachers provide a greater degree of substantive affirmations to Black students relative to White students for the same mathematical solutions (β = 0.22, p=.10).

Table 4. Results from SEM and Multiple Regression Models

	Judgments About the Quality of Student Thinking		Expectation of Task Difficulty		Use of Student Thinking (UST)		Supporting Student Sensemaking (SSS)		Affirmation (TA)	
Black Classroom	0.07 (0.18)	0.09 (0.19)	0.16** (0.05)	0.11* (0.06)	0.17 (0.12)	0.13 (0.12)	0.24* (0.11)	0.20† (0.12)	0.26* (0.12)	0.21† (0.13)
Latinx/e Classroom	0.07 (0.18)	0.07 (0.19)	0.13* (0.05)	0.14* (0.06)	0.12 (0.12)	0.12 (0.12)	0.23† (0.12)	0.17 (0.12)	0.01 (0.12)	-0.03 (0.13)
Woman		0.18 (0.23)		-0.12† (0.07)		-0.04 (0.15)		0.31* (0.14)		0.02 (0.15)
Years Teaching		0.00 (0.05)		-0.05** (0.01)		0.07* (0.03)		0.05 (0.03)		-0.02 (0.03)
Grade level		-0.04 (0.09)		0.14*** (0.03)		-0.04 (0.06)		0.01 (0.06)		0.05 (0.06)
Black Teacher		0.20 (0.31)		0.13 (0.09)		-0.36† (0.20)		-0.35† (0.19)		-0.30 (0.21)
Latinx/e Teacher		-0.01 (0.35)		0.00 (0.11)		0.26 (0.23)		0.38† (0.22)		0.11 (0.24)
AAPI Teacher		-0.30 (0.47)		-0.20 (0.13)		0.15 (0.27)		0.51† (0.27)		0.00 (0.29)
Native Am. Teacher		0.57 (0.77)		-0.12 (0.29)		-1.37** (0.50)		-0.58 (0.48)		-0.24 (0.53)
Multiracial Teacher		-0.07 (0.31)		-0.07 (0.09)		-0.04 (0.20)		-0.22 (0.20)		-0.21 (0.21)
Urbanicity		0.49 (0.45)		-0.16 (0.14)		0.14 (0.29)		-0.10 (0.28)		0.09 (0.31)
% White students		-0.29 (0.47)		-0.03 (0.13)		0.02 (0.30)		0.24 (0.29)		-0.15 (0.32)
Metro Area SES		0.07 (0.18)		0.04 (0.05)		0.03 (0.11)		0.10 (0.11)		0.02 (0.12)
Math Achievement		0.20 (0.44)		-0.37** (0.13)		-0.11 (0.29)		-0.30 (0.28)		-0.08 (0.30)
					$R^2 =$ 0.01	$R^2 =$ 0.07	$R^2 =$ 0.02	$R^2 =$ 0.09	$R^2 =$ 0.02	$R^2 =$ 0.04

****p < .001, **p<0.01, *p<0.05, †<0.10*
Note: Coefficients are shown with standard errors in parentheses. *N*=312 for all unconditional models and *N*=307 for all conditional models due to five observations missing SEDA covariates. For our two scale measures (judgments and expectations), we interpret coefficients on a 1-5 scale. For our coded responses (UST, SSS, and TA), we interpret coefficients as the estimated difference in teacher levels of each standardized latent construct, θ, for a one-unit difference in the predictor.

DISCUSSION

In this study, we replicate and extend the literature about biases in teacher expectations to show that teachers' expectations of math task difficulty for the typical student, but not their judgments about the quality of students' mathematical thinking, may be biased against minoritized students of color. We further expand the evidence base on biases in education to include verbal teacher responsiveness to correct and sometimes non-standard student solutions, finding study participants were more likely to affirm and support the sense-making of target students in the Black condition. We situate each of our findings in the existing literature below, then consider implications for research and practice.

Prior studies have demonstrated biases in teacher judgments about students' math capabilities when reviewing partially correct solutions, with math ability ratings for Black and Latinx/e girls being the lowest compared to boys and White students (Copur-Gencturk et al., 2019). However, we did not find significant differences by condition in teachers' judgments about the quality of students' mathematical thinking. One reason might be that even though some of our vignettes featured non-standard methods, all contained only correct student work, with some of those solutions demonstrating substantial mathematical reasoning. These findings suggest that stereotype-disconfirming information, such as correct and even sophisticated mathematical solutions, may diminish the potential for biases to emerge (Fiske & Neuburg, 1990; Pendry & Macrae, 1994).

Next, our findings show teachers considered math tasks more difficult for the typical student when presented with 'classrooms' primarily composed of Black and Latinx/e students as opposed to White students. Because teacher expectations relate to curricular choices (Aydin & Ok, 2022), this finding may be consequential for the curriculum students receive. For instance, expectations about task difficulty may underlie the biases against Black students in teachers' curricular selection in mathematics found in Pollard (2022). As our findings showed teachers expect greater difficulty for Black and Latinx/e learners based on the same mathematical content, our study raises questions about how teachers attribute those difficulties and what kinds of instructional experiences follow, with important implications for racial equity in education.

This study also extends prior findings on biases in instruction to the domain of ambitious math teaching. Our findings suggest teachers may be more likely to support student sensemaking in their next instructional steps for Black students relative to White students. In similar work, Pollard (2022) also showed that teachers were more likely to select student-centered instructional approaches for Black students compared to White students. That teachers prioritized student-centered instructional approaches in the Black condition may run counter to hypotheses about racial biases, suggesting teachers may be taking more asset-based views of Black students. Teachers' demonstration of student-centered approaches for Black learners might also reflect an orientation towards culturally responsive and interactive teaching approaches for these students. Given the pronounced mathematics instructional reforms of the past two decades emphasizing student-centered approaches, particularly in urban districts serving BIPOC students (Anderson, 2014), it could be the case that teachers may be supporting Black students' math learning in unique ways.

Conversely, social desirability may shape teachers' responses. So might teacher attendance at anti-bias professional development, which started taking place more frequently in the wake of George Floyd's murder by police officers in May 2020, about 18 months before this study launched. Examining whether teachers' responses to our survey translate into classroom instruction would help ascertain the extent to which pro-Black biases in supporting student sense-making occurs in practice, and whether they carry any unintended instructional consequences for students.

When examining teacher responsiveness to students' mathematical thinking, we find emergent evidence to suggest teachers may provide more positive and substantive affirmations for Black students relative to White students for the same mathematical content. These findings extend prior research by Harber et al. (2012) and Kleen & Glock (2018)—which found teachers demonstrate positive feedback bias toward Black, Latinx/e, and ethnic minority students—to the domain of ambitious math teaching. These findings align with the shifting standards theory, which states that stereotypes about social groups inform standards for evaluating members of that group, and those standards shift and differ across groups (Biernat et al., 1991; Biernat, 1995). Thus, this finding may result from teachers showing a positive bias for minoritized students who excel academically. In similar work, Kleen & Glock (2018) found teachers judged high-performing ethnic minority students more favorably than high-performing ethnic majority students.

Fiske & Neuburg's (1990) dual-process theory for arriving at social judgments may help to interpret our seemingly contradictory findings. As our intervention presented teachers in all conditions with a student's proficient mathematical thinking and primed them to respond, teachers may have been confronted with information that contradicted activated stereotypes about learners. Our results showed teachers expected a greater degree of difficulty for students of color, yet they evaluated the quality of and used students' mathematical thinking comparably across conditions. Here, teacher perceptions about student proficiency may have contradicted any stereotypes, mitigating the potential for bias. When presented with information about a person that contradicts a stereotype, an information-integrating strategy is employed whereby individual attributes are processed in an effortful, conscious manner to arrive at a judgment (Fiske & Neuburg, 1990). Such effortful and controlled processing may have worked to counteract perceived stereotypes (Pendry & Macrae, 1994), resulting in relatively equitable evaluations for and use of student thinking across groups.

LIMITATIONS

We reflect on several limitations to the present study. To start, we may have inadvertently intervened on what we intended to measure: it could be the case that we did not observe biases in some domains of responsive math instruction because we prompted all teachers in our study to respond to the student's mathematical thinking. An understanding of teachers' responses in the classroom also includes *not* responding (Kang, 2022). It remains to be observed whether educators would have chosen to respond to the student's mathematical ideas had they not been universally prompted to do so. Prompting teachers to take up student thinking could have played a leveling role in student opportunities to learn.

Another limitation concerns the racial make-up of our sample and not having enough power to enable an examination of interaction effects between teacher and student race-ethnicity. Our sample consisted primarily of White educators, and White teachers may demonstrate social desirability bias in efforts to maintain a positive self-image and not appear biased or racist (Crosby & Monin 2007; Plant & Devine, 1998). Moreover, there is ample literature on the positive effects of racial-ethnic congruence or "match" between teachers and students (Gershenson et al., 2016; Redding, 2019). Future work investigating racial biases in ambitious math instruction should include a racially diverse sample of teachers and attend to the phenomenon of racial-ethnic match between teacher and student.

A final limitation stems from the study's design, which precluded us from two important lines of inquiry. First, due to sample size limitations, we could not take an intersectional approach that examines interactions between student gender and race-ethnicity. The second limitation from the study design concerns the content of the vignettes. All student answers and solutions featured in the vignettes presented were mathematically correct, with some even demonstrating students reasoning through non-standard problem-solving strategies. This context may have made it clear to teachers that students across groups were equally proficient in the particular mathematics topics represented in the vignettes. This is further supported by the fact that we did not detect any differences in teachers' use and judgments about the quality of student thinking. Such correct solutions may have provided stereotype-disconfirming evidence to teachers, reducing the potential influence of racial biases. Student demonstration of mathematical proficiency is consequential for the instructional character of a classroom. Indeed, prior research has shown teachers to employ ambitious teaching practices in math less frequently the more students they had performing below grade level (Gottfried et al., 2023). Future work in this area should examine the role of student math proficiency in teachers' racial-ethnic biases in this domain.

CONCLUSIONS

Youth in U.S. schools are more racially and ethnically diverse than they have ever been and are projected to be majority-minority by 2050 (Rivas-Drake & Umaña-Taylor, 2019), meanwhile the teaching profession remains predominantly White (National Center for Education Statistics, 2023). Such demographic change is not inconsequential, as White educators in the U.S. have been found to carry negative stereotypes and social distance from minoritized groups (Quinn & Stewart, 2019), and U.S. counties with higher aggregate teacher racial bias tend to show greater achievement disparities between Black and White students (Chin et al., 2020). Ambitious math teaching is characteristic of effective learning environments for children, so much so that there have been concerted efforts to rehaul mathematics instruction in this way, particularly in urban schools serving Black and Latinx/e communities (Anderson, 2014). Our findings indicate the wide scale implementation of math instructional reforms absent a consideration of race could undermine goals for equitable and just math teaching. Knowledge about the specific instructional sub-domains and circumstances subject to biases better equips teacher educators and practitioners to intervene on them, for example, through practice-in-context (Mancenido et al., 2025) and coaching (Cohen et al., 2024) to develop pre-service teacher capacity to deliver equitable mathematics learning experiences. The findings from this study have implications for pre-service teacher education and professional development to support educator capacity for delivering rigorous, equitable, and just math learning experiences across differences and diverse student groups.

REFERENCES

Alexander, R. J. (2008). Towards dialogic teaching: Rethinking classroom talk (4th ed.). Dialogos.
Anderson, C. R. (2014). Place matters: Mathematics education reform in urban schools. *Journal of Urban Mathematics Education*, *7*(1), 7. https://doi.org/10.21423/jume-v7i1a231
Anderson-Clark, T. N., Green, R. J., & Henley, T. B. (2008). The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology, 27*(1), 94–99.
Authors (2024). Under review
Aydin, Ö., & Ok, A. (2022). A systematic review on teacher's expectations and classroom behaviors. *International Journal of Curriculum and Instructional Studies, 12*(1), 247-274.
Ball, D. (2018). "Presidential Address," American Educational Research Association Annual Conference, New York.
Biernat, M. (1995). The shifting standards model: Implications of stereotype accuracy for social judgment. In Y. T. Lee, L. J. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 87–114). American Psychological Association.
Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*, 485–499. https://doi.org/10.1037/0022-3514.60.4.485
Bonefeld, M., Dickhäuser, O. & Karst, K. (2020). Do preservice teachers' judgments and judgment accuracy depend on students' characteristics? The effect of gender and immigration background. *Social Psychology of Education, 23*(2), 189–216. https://doi-org.ezp-prod1.hul.harvard.edu/10.1007/s11218-019-09533-2
Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal, 113*(1), 76–104. https://doi.org/10.1086/666387
Boston, M. D., & Wilhelm, A. G. (2017). Middle school mathematics instruction in instructionally focused urban districts. *Urban Education, 52*(7), 829-861. https://doi.org/10.1177/0042085915574528
Brophy, J. E. (1981). Teacher praise: A functional analysis. *Review of educational research, 51*(1), 5-32.
Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. Holt, Rinehart and Winston.
Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 499-531.
Charalambous, C. Y. (2020). Reflecting on the troubling relationship between teacher knowledge and instructional quality and making a case for using an animated teaching simulation to disentangle this relationship. *ZDM Mathematics Education, 52*(2), 219-240. http://dx.doi.org/10.1007/s11858-019-01089-x
Chin, M. J., Quinn, D. M., Dhaliwal, T. K., Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, *49*(8):566-578. https://doi.org/10.3102/0013189X20937240
Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2), 294-343.
Cohen, J., Wong, V. C., Krishnamachari, A., & Erickson, S. (2024). Experimental evidence on the robustness of coaching supports in teacher education. *Educational Researcher, 53*(1), 19-35. https://doi.org/10.3102/0013189X231198827
Copur-Gencturk, Y., Cimpian, J. R., Lubienski, S. T., & Thacker, I. (2019). Teachers' bias against the mathematical ability of female, Black, and Hispanic students. *Educational Researcher*, *49*(1), 30-43. https://doi.org/10.3102/0013189X19890577
Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2022). Teacher bias in the virtual classroom. *Computers & Education*, *191*, 104627. https://doi.org/10.1016/j.compedu.2022.104627
Copur-Gencturk, Y., Thacker, I., & Cimpian, J. R. (2023). Teachers' race and gender biases and the moderating effects of their beliefs and dispositions. *International Journal of STEM Education*, *10*(1), 31. https://doi.org/10.1186/s40594-023-00420-z
Crosby, J. R., & Monin, B. (2007). Failure to warn: How student race affects warnings of potential academic difficulty. *Journal of Experimental Social Psychology, 43*(4), 663–670. https://doi.org/10.1016/j.jesp.2006.06.007
Cross, W. E. Jr., & Cross, T. B. (2008). Theory, research, and models. In Quintana, S. M. & McKown, C (Eds.), *Handbook of race, racism, and the developing child* (pp. 154–181). Wiley.

Dovidio, J. F. (2001). On the Nature of Contemporary Prejudice: The Third Wave. *Journal of Social Issues*, *57*(4), 829–849. https://doi.org/10.1111/0022-4537.00244

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Harcourt Brace Jovanovich College Publishers.

Ewing, E. L. (2020, July 2). I'm a Black scholar who studies race. here's why I capitalize 'White.' *ZORA*. https://zora.medium.com/im-a-black-scholar-who-studies-race-here-s-why-i-capitalize-white-f94883aa2dd3

Fahle, E. M., Chavez, B., Kalogrides, D., Shear, B. R., Reardon, S. F., & Ho, A. D. (2021). *Stanford Education Data Archive: Technical Documentation (Version 4.1).* Retrieved from http://purl.stanford.edu/db586ns4974.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Academic Press. https://doi.org/10.1016/S0065-2601(08)60317-2

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal, 91*(1), 29–42. https://doi.org/10.1353/hsj.2007.0022

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209–224. https://doi.org/10.1016/j.econedurev.2016.03.002

Gottfried, M., Fletcher, T., & Comstock, M. (2023). Teaching mathematics in kindergarten: How does instruction differ by classroom ability? *Teachers College Record, 125*(10), 102–130. https://doi.org/10.1177/01614681231223343

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. Psychological Review, *102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2014). Supplemental Material for Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies. *Journal of Personality and Social Psychology*, supp. https://doi.org/10.1037/a0032734.supp

Gutiérrez, R. (2013). The Sociopolitical Turn in Mathematics Education. *Journal for Research in Mathematics Education*, *44*(1), 37–68. https://doi.org/10.5951/jresematheduc.44.1.0037

Gutiérrez, R. (2012). Context matters: How should we conceptualize equity in mathematics education? In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in discourse for mathematics education: Theories, practices, and policies* (pp. 17–33). Springer. https://doi.org/10.1007/978-94-007-2813-4_2

Harber, K. D., Gorman, J. L., Gengaro, F. P., Butisingh, S., Tsang, W., & Ouellette, R. (2012). Students' race and teachers' social support affect the positive feedback bias in public schools. *Journal of Educational Psychology*, *104*(4), 1149–1161. https://doi.org/10.1037/a0028110

Herbel-Eisenmann, B. A., Steele, M. D., & Cirillo, M. (2013). (Developing) teacher discourse moves: A framework for professional development. *Mathematics Teacher Educator, 1*(2), 181-196.

Herbst, P., Chazan, D., Chen, C. L., Chieu, V. M., & Weiss, M. (2011). Using comics-based representations of teaching, and technology, to bring practice to teacher education courses. *ZDM Mathematics Education, 43,* 91-103. https://doi.org/10.1007/s11858-010-0290-5

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237–271. https://doi.org/10.1146/annurev.psych.47.1.237

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality & Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives*. Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Irvine, J. J. (2010). Foreword. In Milner's H. R. (Ed.), *Culture, curriculum, and identity in education* (pp. xi-xvi). Palgrave Macmillan

Jacobs, V. R., & Empson, S. B. (2016). Responding to children's mathematical thinking in the moment: An emerging framework of teaching moves. *ZDM Mathematics Education*, *48*, 185-197. https://doi.org/10.1007/s11858-015-0717-0

Jencks, C. & Phillips, M. (Eds.). (2011). *The Black-White test score gap*. Brookings Institution Press,

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9,* 131-155. http://dx.doi.org/10.1207/ s15327957pspr0902_3.
Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), Advances in experimental social psychology, Vol. 28, pp. 281–388). Academic Press. https://doi.org/10.1016/S0065-2601(08)60240-3
Kang, H. (2022). Teacher responsiveness that promotes equity in secondary science classrooms. *Cognition and Instruction*, *40*(2), 206-232.
Kleen, H., & Glock, S. (2018). A further look into ethnicity: The impact of stereotypical expectations on teachers' judgments of female ethnic minority students. *Social Psychology of Education, 21*(4), 759–773. https://doi.org/10.1007/s11218-018-9451-0
Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling (4th ed.). Guilford Press.
Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, *35*(7), 3–12. https://doi.org/10.3102/0013189X035007003
Lampert, M. (1990). Connecting inventions with conventions. In L. P. Steffe & T. Wood (Eds.), Transforming children's mathematics education (pp. 253-265). Lawrence Erlbaum Associates, Inc.
Lampert, M. (2001). *Teaching problems and the problems of teaching*. Yale University Press.
Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education, 64*(3), 226–243. https://doi.org/ 10.1177/0022487112473837
Laws, M. (2020). Why we capitalize 'Black' (and not 'white'). Columbia Journalism Review. Retrieved August 9, 2024, from https://www.cjr.org/analysis/capital-b-black-styleguide.php
Lightfoot, S. L. (1978). *Worlds apart: Relationships between families and schools*. Basic Books
Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education, 60*, 245–256. https://doi.org/10.1177/0004944116664618
Mancenido, Z., Hill, H., Garcia Coppersmith, J., Carter, H., Pollard, C., & Monschauer, C. (2025). Practice-based teacher education pedagogies improve responsiveness: Evidence from a lab experiment. *Journal of Research on Educational Effectiveness*. (forthcoming)
Marshall, Y. (2020) (In)articulate while Black—AAIHS. *Black Perspectives.* Retrieved August 11, 2024, from https://www.aaihs.org/inarticulate-while-black/
Martin, D. B. (2009). Liberating the production of knowledge about African American children and mathematics. In D. B. Martin (Ed.), *Mathematics teaching, learning, and liberation in the lives of Black children* (pp. 3–36). Routledge/Taylor & Francis Group.
Milner, H. R. (2010). *Start where you are, but don't stay there: understanding diversity, opportunity gaps, and teaching in today's classrooms*. Harvard Education Press.
Milner, H. R. (2012). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies, 43*(6), 693-718. https://doi-org.ezp-prod1.hul.harvard.edu/10.1177/0021934712442539
Nasir, N. S., Hand, V., & Taylor, E. V. (2008). Culture and mathematics in school: Boundaries between "cultural" and "domain" knowledge in the mathematics classroom and beyond. *Review of Research in Education*, *32*, 187–240.
National Center for Education Statistics. (2023). *Characteristics of public school teachers*. Condition of education. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/clr
National Center for Education Statistics. (2019). *Trends in International Mathematics and Science Study (TIMSS).* National Center for Education Statistics.
National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. NCTM.
National Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: ensuring mathematical success for all*. NCTM.
National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics.*
O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussion. In D. Hicks (Ed.), *Discourse, learning, and schooling* (pp. 63-103). Cambridge University Press.

Pendry, L. F., & Macrae, C. N. (1994). Stereotypes and mental life: The case of the motivated but thwarted tactician. *Journal of Experimental Social Psychology, 30*, 303–325. https://doi.org/10.1006/jesp.1994.1015

Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, *42*, 123–140. https://doi.org/10.1016/j.learninstruc.2016.01.010

Pit-ten Cate, I. M., & Glock, S. (2023). A systematic review on teachers' stereotypical beliefs and expectations: Effects of the intersectionality of students' gender and ethnicity. In P. W. St. J. Watson, C. M. Rubie-Davies, & B. Ertl (Eds.), *Routledge international handbook of gender beliefs, stereotype threat and teacher expectations* (1st ed., pp. 238–250). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781003275763-24

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*(3), 811–832. https://doi.org/10.1037/0022-3514.75.3.811

Pollard, C. (2022). *Teaching, learning, and race: Toward the identification of mechanisms underlying own- and other-race teacher effects.* [Doctoral dissertation, Harvard University Graduate School of Arts and Sciences]

Quinn, D. M. (2020a). Experimental effects of "achievement gap" news reporting on viewers' racial stereotypes, inequality explanations, and inequality prioritization. *Educational Researcher*, *49*(7), 482–492. https://doi.org/10.3102/0013189X20932469

Quinn, D. M. (2020b). Experimental evidence on teachers' racial bias in student evaluation: the role of grading scales. *Educational Evaluation and Policy Analysis*, *42*(3), 375–392. https://doi.org/10.3102/0162373720932188

Quinn, D. M., & Stewart, A. M. (2019). Examining the racial attitudes of white pre-K–12 educators. *The Elementary School Journal*, *120*(2), 272–299. https://doi.org/10.1086/705899

Reardon, S. F., Cimpian, J., & Weathers, E. S. (2015). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. F. Ladd, & M. E. Goertz (Eds.), *Handbook of Research in Education Finance and Policy* (2nd ed.). Routledge.

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, *89*(4), 499–535. https://doi.org/10.3102/0034654319853545

Richards, J., & Robertson, A. D. (2015). A review of the research on responsive teaching in science and mathematics. In *Responsive teaching in science and mathematics* (1st ed., pp. 36–55). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315689302-2

Rivas-Drake, D., & Umaña-Taylor, A. (2019). *Below the surface: Talking with teens about race, ethnicity, and identity*. Princeton University Press.

Rubie-Davies, C. M., Flint, A., & McDonald, L.G. (2012). Teacher beliefs, teacher characteristics and school contextual factors: What are the relationships? *British Journal of Educational Psychology, 82*(2), 270–288. https://doi.org/10.1111/j.2044-8279.2011.02025.x

Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology, 76*(3)*,* 429–444. https://doi.org/10.1348/000709905X53589

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 34*, 1–97.

Schneider, D. J. (2004). *The psychology of stereotyping*. Guilford Press.

Schoenfeld, A., Fink, H., Zuñiga-Ruiz, S., Huang, S., Wei, X., & Chirinda, B. (2023). *Helping students become powerful mathematics thinkers: Case studies of teaching for robust understanding*. https://doi.org/10.4324/9781003375197

Shaughnessy, M., & Boerst, T. A. (2018). Uncovering the skills that preservice teachers bring to teacher education: The practice of eliciting a student's thinking. *Journal of Teacher Education, 69*(1), 40-55. https://doi.org/10.1177/00224871177025

Smylie, M., & Wenzel, S. (2006). *Promoting instructional improvement: A strategic human resource management perspective.* Consortium on Chicago School Research.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, *10*(4), 313–340. https://doi.org/10.1080/10986060802229675

Sun, K. L., & Ruef, J. L. (2023). Examining and conceptualizing the relationship between teacher praise and the co-construction of mathematical competence in classrooms. *The Journal of Mathematical Behavior, 71*(3), 101065. https://doi.org/10.1016/j.jmathb.2023.101065

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–103.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, *99*(2), 253–273. https://doi.org/10.1037/0022-0663.99.2.253

The New Teacher Project (TNTP) (2018). *The opportunity myth*. New York, NY: Author. Retrieved September 20, 2018, from https://tntp.org/assets/documents/TNTP_The-Opportunity-Myth_Web.pdf

Trejos, J. (2024). What's in a "Latinx?": Considerations when utilizing pan-ethnic identifiers of Latinidad. *Columbia Social Work Review, 22(*1), 28-48. https://doi.org/10.52214/cswr.v22i1.12235

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*(2), 497–527. https://doi.org/10.3102/0002831209353594

Warikoo, N., Sinclair, S., Fei, J., & Jacoby-Senghor, D. (2016). Examining racial bias in education: A new approach. *Educational Researcher, 45(9),* 508–514. https://doi.org/10.3102/0013189X16683408

Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*. Harvard University Press. https://doi.org/10.4159/9780674045040

APPENDIX

Coding Scheme for Teacher Responsiveness Measures

Eliciting Student Thinking (EST)			
N/A	0 - No eliciting of student thinking.	1 - Some elicitation of student thinking	2 - Strong elicitation of student thinking
No teacher utterance or teacher utterance is not responsive to prompt.	Teacher makes a statement, whether approval or correction. No expectation of student response. Asks a question to the whole of the class / not the initial respondent (e.g. "*What do others think*?; "*Did anyone else have a different way?")*	Teacher poses a simple proforma clarifying or probing question that does not draw specifically on what the student has shared (e.g. *"How did you get that?"; "Can you explain to everyone what you mean?")*	Teacher poses a question to clarify or probe a specific idea that the student has shared.

Using Student Thinking (UST)			
N/A	0 - No or minimal use of student thinking	1 - Some use of student thinking	2 - Strong use of student thinking
No teacher utterance or teacher utterance is not responsive to prompt.	Response is plausible in the given situation (e.g. *"Ok, nice!"*). Teacher responds in a pro forma way (e.g., acknowledges student response is correct or incorrect; thanks or praises the student but does not respond to their mathematical ideas; asks for another student answer to the problem; provides direct instruction about the problem to the class).	Teacher response goes beyond pro forma to feature some use of student ideas; focus stays at least for a moment on the student(s) but it does not rise to strong use (e.g., brief restatement of student method; brief restatement followed by teacher direct instruction; asking students to repeat their answers; asks class whether student is correct).	Teacher weaves student ideas into the development of the mathematics (e.g., compares student responses; adds mathematical emphasis to student method or solution; fills in mathematical details that were missing from student response; asks student "why" question; asks class "why" question based on a student method; asks "*How do you know?")*

Supporting Student Sensemaking (SSS)		
N/A	0 - No support of student sensemaking	1 - Support of student sensemaking
No teacher utterance or teacher utterance is not responsive to prompt	Teacher does not invite student input or activity related to the original student statement. The teacher instead moves on to direct instruction, simple approval, or asks the class to move on to a **different* problem or task. The nature of student input is such that students are asked to answer a question for which the teacher already has a clear and strict answer in mind. * *Different* means the new task is unrelated/ unresponsive to where students might be at, e.g. "*Let's do another problem, class*")	Teacher keeps the student idea in play somehow and/ or attends to student thinking in the classroom. The teacher "has an ear to the ground" and seeks student input, ideas, and reasoning OR assigns an additional task related to the original student idea or related to student thinking "on the ground". The nature of student input is such that students are given the opportunity to engage in their own sense-making. E.g. "*Who else agrees with this answer? Did anyone solve it a different way? Turn and talk to your partner about X's answer."*

Teacher Affirmation of Student Math Ideas (TA)			
N/A	0 - No teacher affirmation of student ideas/ work	1 - Generic teacher affirmation of student ideas/ work	2 - *Specific* teacher affirmation of student ideas/ work
No teacher utterance or teacher utterance is not responsive to prompt	Teacher provides <u>no praise</u>. Does not position students as competent and does not affirm what they did that was mathematically productive. Teacher does not affirm the student work or provide any type of praise.	Teacher assigns praise to the student in the form of positive content or tone, <u>but it is non-specific/ generic</u> (e.g. *"Good job!"* without specifying the mathematical idea or specific thing a student did). Teacher may only address the correctness of the answer or <u>student engagement</u>, but not the mathematical idea.	Describes the extent to which the teacher demonstrates the ability to position students as competent with affirmations <u>specifying what they did that was mathematically productive</u>. Teacher actively affirms the specific student work (e.g., calling out what is correct in the student work, noting a strength in student response).