Race-gender D/discourses in mathematics education: (Re)-producing inequitable participation patterns across a diverse, instructionally-advanced urban district

Daniel Reinholz & Annie Wilhelm

This paper uses quantitative analytics to study talk-based participation in 100 mathematics classrooms across one racially diverse urban school district in the USA. Using the EQUIP observation tool and hierarchical linear modeling, we characterize the quantity and quality of participation for students across 3025 coded turns, by race and gender. We found that in general, boys participated significantly more than girls. We also found that Latinx and Asian/Pacific Islander students had significantly fewer turns than Black and White students. To interpret these findings in context, we analyzed interviews from 29 district leaders using a poststructural framework organized around D/discourse.

Keywords: d/Discourse, poststructural theory, race-gender subgroups

Introduction

In 2022, white supremacy is on full display. From the murders of unarmed Black people, immigrant Latinx children locked in cages, violent attacks on Asian Americans, and the decimation of Indigenous communities from COVID-19, historical legacies of oppression have come to the fore. This sociopolitical context is also consequential for understanding inequities in urban mathematics education. White supremacy and patriarchy prevalent in society writ large also permeate mathematics classrooms (Lubienski & Ganley, 2017; Martin et al., 2017).

The connection between the mathematics classroom and society can be understood from a poststructuralist perspective (Foucault, 1977). The same problematic cultural representations (i.e., Discourses) that percolate in society are also ready to be deployed and position students as capable at mathematics or not (Shah, 2017). Discourses manifest, for example, by constraining student opportunities for participation, learning, and identity development.

Our study attends to student participation across a sample of 100 classrooms in a single, racially diverse "urban emergent" (Milner, 2012, p.559) district. We focus on middle school mathematics, because it is a critical juncture in mathematics learning (Faulkner et al., 2014). Using the classroom observation tool EQUIP (Reinholz & Shah, 2018), we addressed the following research questions: *1) What are district leaders' Discourses about students in relation to mathematics? And, 2) How does participation in whole-class mathematical discourse differ by gender and race across the district?*

Background

We situate our study within the context of urban mathematics education. Historically, "urban" has been used as coded language to disparage students of color in racially diverse schools (cf. Milner, 2012), and indeed, urban has been used a proxy for deficit-oriented achievement gap rhetoric in mathematics education writ large (Gutiérrez, 2008). Such rhetoric frames racially minoritized students in urban schools as a problem to be solved, when in reality, colorblind racist rhetoric (re)produces these inequities (Berry et al., 2014).

Recent efforts reclaim urban mathematics education as a political project to disrupt oppression in mathematics education (Larnell & Martin, 2021). This work constructs counternarratives about racially minoritized mathematics learners and their communities (Jett, 2019; Stinson, 2013). Powerful counterstories are not purely conceptual, but actually influence policy and practice (Goldin et al., 2021). For example, anti-deficit approaches to interpreting achievement gaps (e.g., in terms of opportunity gaps), support conversations about de-tracking, culturally relevant pedagogy, and challenge teacher biases (Flores & Gunzenhauser, 2021; Taylor et al., 2021). In the present study, we use the related concepts of D/discourse[1] to understand differences in opportunities to participate, and thus learn, at the scale of a district.

Poststructuralism and Discourse

A central object of study in poststructuralism is a Discourse, which is a collection of symbols, signs, artifacts, and other cultural representations that work together to constitute the social world. Discourses consist of *subject positions* occupied by individuals, which constrain their possible actions within the Discourse. Discourses exert power, as they guide and constrain how people behave (Foucault, 1977).

---

[1] We use capital "D" Discourse to denote broader societal constructions, whereas lowercase "d" discourse refers to mathematics discussions in local classrooms.

Consider the social construct gender. Gender divides the world into different subject positions (e.g., man, woman), and exerts power over individuals by dictating acceptable behaviors, from the clothes one wears to their career aspirations. Of course, Discourses are not immutable. Binary gender Discourses have long been contested, and sometimes, rejected entirely. Moreover, particular positions can be taken up in different ways (e.g., there are different types of women), and as individuals create and inhabit new subject positions (e.g., a non-binary person), they can shift the overall Discourse. In any context, multiple Discourses exist, which may intersect gender and many other identities, creating a vast array of potential subject positions.

Race-Gender Discourses in Urban Mathematics Education

In mathematics, students are often divided into categories like "those who can do math" and "those who cannot" (Wai, 2012). Similarly, racial Discourses like "Black boys are a problem" (Gholson, 2016) or "Latinas will be quiet and compliant" (Shah, Herbel-Eisenmann, et al., 2020), shape how students are seen by their teachers (and other students). How students are positioned within a given classroom is connected to the prevalence of these broader Discourses at school and institutional levels (Sengupta-Irving, 2021). This positioning is also related to the presence of broader institutional structures, such as tracking. Especially when judgments are grounded in false racial narratives, tracking into different ability groups becomes a powerful tool for sustaining racism (Oakes, 1990).

Understanding how Discourses operate requires an intersectional approach (Crenshaw, 1990). Intersectionality highlights how systems of oppression are interlocking and mutually constitutive. For example, racism and ableism intertwine to overidentify and underserve racially minoritized students in special education (Annamma et al., 2019). Similarly, Black women face forms of gender oppression that white women do not (Gholson, 2016).

Discourses like the "achievement gap" negatively position all racially minoritized students, especially Black, Latinx, and Indigenous students (Gutiérrez, 2008). Material consequences of such Discourses manifest through low expectations, limited opportunities to participate, microaggressions, stereotypes, and stereotype threats (Jackson, 2009; Martin et al., 2017). For example, when Black children are framed as deficient, teachers describe lower expectations for those children (Martin, 2012). Gendered Discourses operate through similar mechanisms that impact women and other gender minorities (Lubienski & Ganley, 2017; Voigt & Reinholz, 2020).

Positioned as the "default" in mathematics, nondisabled cisgender white men are often positioned as a standard by which to compare others (Battey & Leyva, 2016; Reinholz, 2022). Racial and gender Discourses intersect, for example, when Black boys are positioned as aggressive and violent (Leath et al., 2019), and by comparison, Black girls are largely invisibilized (Carter Andrews et al., 2019). Similarly, Discourses that Latinx students do not care about school or cannot speak English combine with Discourses around femininity to invisibilize Latinx girls (Shah, Herbel-Eisenmann, et al., 2020). Indigenous students have received less attention in mathematics education and are often positioned as a part of the collective "other," which is another form of invisibilization.

In contrast, students of Asian ancestry have been positioned as a model minority (Shah, 2017). Nonetheless, narratives such as "Asians are good at math" are dehumanizing and framing Asian students as "always successful" ignores variations in culture, language practices, immigration status, and socioeconomic conditions. Research shows that when Asian populations are disaggregated, these variations account for significant differences in outcomes (Pang et al., 2011).

Visibility and discourse Practices

In contrast to Discourses (capital D) that broadly position students, discourses (lowercase d) focus on how students engage in classroom discussions. Prior work has detailed the subtle ways that discourse practices position students (e.g., Engle et al., 2014), and similarly, how instructors can adapt their own practices to shift power hierarchies (Cohen & Lotan, 2014). Research shows that these two processes are interrelated, for example, as broader Discourses about Black learners influence how their teachers position them within classrooms

(Wilkes & Ball, 2020). Thus, to understand the learning opportunities afforded to particular groups of learners, tracing broader Discourses should be one productive way to contextualize what happens locally (Sengupta-Irving, 2021).

Participation in mathematical discourse is important because it supports learning (Ing et al., 2015; Reinholz et al., 2022) and identity development (Nasir et al., 2008). In short, students who have more opportunities to participate and be seen as competent, have more opportunities to succeed in mathematics (Reinholz & Shah, 2018). Nonetheless, research shows that these opportunities are not fairly distributed across race and gender (Ernest et al., 2019; Reinholz et al., 2022; Shah, Christensen, et al., 2020). From a poststructuralist perspective, student positioning in broader Discourses mediates their opportunities to participate in local discourses (i.e., negative race-gender stereotypes result in limited opportunities to participate).

Significance of the Study

This study makes a major methodological contribution due to its scale. Prior EQUIP-based work has documented how inequities arise in local contexts by considering samples of a few classrooms (e.g., Reinholz et al., 2019; Reinholz & Shah, 2018; Shah, Christensen, et al., 2020). This is similar to other research on classroom discourse, which typically focuses on only one or a few classrooms (e.g., Engle et al., 2014). Scaling local approaches to hundreds of classrooms is nontrivial, but necessary to move educational policy. While we recognize tradeoffs are inherent, this work pushes the field forward, exploring D/discourses across an entire district.

Second, our study looks at phenomena disaggregated across different racial groups. We carefully attend to race-gender subgroups within student participation. Although a body of research documents how systems of oppression are interlocking, such systems are often studied separately in practice. This occurs for a variety of reasons, including limitations in data collection, research design, or analysis. Thus, our results allow us to move beyond singular stories of racial or gender marginalization, to understand race-gender Discourses together.

Method

We performed a secondary analysis of data from the MIST project, which conducted a wide array of interviews and collected a substantial amount of classroom video across four school districts. Using the EQUIP tool, we extend the original study to attend to racial and gender inequity in one school district.

Context and Participants

MIST Project

The Middle School Mathematics and the Institutional Setting of Teaching (MIST) project was a 4-year longitudinal study of middle school mathematics teaching in four large urban school districts (Cobb et al., 2020). The four districts were selected because of their emphasis on improving the quality of teaching and learning in mathematics classrooms. The MIST project sought to understand what was needed to improve the quality of middle grades mathematics teaching and thus student achievement in large, urban school districts.

The District and Sample

For this analysis, we focused on District A, because it was the most racially and ethnically diverse district in the MIST project sample (see Figure 1, and supplementary materials Table B), and it was ahead of the other districts in its implementation of ambitious instructional practices (Boston & Wilhelm, 2017). This district can be classified as an *urban emergent* district (Milner, 2012, p.559), given its location in a large city with fewer than one million people.

At the time of the study (2007-2011), District A served around 35,000 students, with approximately the following breakdown by race: 38% Black, 30% White, 18% Latinx, 8% Asian/Pacific Islander (API), 5% Native American, and 1% other races[2]. Within each of these racial categories, the percentage of boys

[2] Our descriptions are limited by the types of demographic data that the districts provided for their respective states. For example, District A identifies students who have emigrated from countries in Africa as African American, even though they have unique histories of education and otherwise.

and girls differed by at most 1%. Individual schools varied significantly in their racial breakdown (see Figure 1). Two schools (Schools 2 and 3) were over 50% White, 5 schools (Schools 1, 5, 7, 9, and 10) near or over 50% Black, one school (School 4) at about 50% Latinx, and the other two schools (Schools 6 and 8) had more evenly distributed racial breakdowns across the three most represented categories.

Schools generally ranged from 1% to 29% of students classified as English Language Learners (ELLs).[3] While there were nearly 80 languages represented in the district, ELLs primarily belonged to three groups: Latinx (16%), Somali (6%), and Hmong (5%) communities. The two schools with lowest percentages of ELLs had the greatest percentage of White students (Schools 2 and 3). The 10 study schools generally served between 400 and 750 students, with the larger schools serving elementary and middle school student populations. All schools other than Schools 2 and 3 were classified as Title 1 schools, which meant that at least 40% of the student population qualified as low-income, and this made the schools eligible for additional state and national funding.

District A was instructionally ahead of the other districts in the MIST study. Most notable was that in about half of observed lessons, tasks were

---

Additionally, there are likely students whose roots are not in Africa but who are nonetheless identified in school as African American (e.g., those who identify as Caribbean or West Indian).

[3] The term ELL was a classification provided by the districts. We recognize this is deficit-oriented language that negatively positions students learning English, compared to emergent multilingual, which better recognizes student strengths.

implemented at high levels, often with high level discussions of student thinking and reasoning, which was significantly more often than in other districts (Boston & Wilhelm, 2017). There are a number of factors that likely contributed to these relatively ambitious instructional practices (Boston & Wilhelm, 2017; Cobb et al., 2020). Teachers in District A were significantly more experienced than teachers in the other districts. The district had a long-standing commitment to mathematics instruction, including over 10 years of experience using Connected Mathematics Project curriculum materials (Lappan et al., 2003) and providing additional professional development in Designing Groupwork (Cohen & Lotan, 2014). Across the district there was also professional development aimed at eliminating Black-White achievement gaps (e.g., culturally responsive teaching, Courageous Conversations; Webster, 2016). Notably, the district was committed to de-tracking and providing access to algebra for all students by eighth grade.

At the beginning of the MIST study, 10 middle schools were chosen to represent the variation in middle schools in the district. Within each school, 2-4 teachers were randomly selected from the 2-5 mathematics teachers in each school, resulting in 30 middle school teachers from the 10 different schools. Teachers participated for all four study years when possible. Teachers who left the study were replaced, so a total of 45 teachers participated, with 2-4 teachers per school at any given time. Those 45 teachers were 58% women and averaged 11.4 years of experience teaching mathematics in their first year in the study.

They were asked to identify with one or more race or ethnicity, with 86.7% identifying as White, 6.7% as Black, 6.7% as API, 2.2% Hispanic or Latinx, and 2.2% as Native American.

Data Sources

To understand broader Discourses in the district, we conducted a secondary analysis of interviews with district leaders (N = 45). In the MIST project, *interview summary forms* were generated for each interview. These forms provided a detailed summary of key points, as well as relevant quotes from interviews. We drew on summaries from 45 interviews spread over 4 years, representing a total of 29 leaders. The interviewed leaders ranged from superintendents to instructional coaches and data analytics specialists. We focused on a specific portion of the interview summary forms that was directly related to our analytic goals:

> *Please describe any categories that the district leader used to characterize students, or groups of students in relation to mathematics. Please record their* exact *categories and the meanings that those categories hold (e.g., "fast students," "geniuses," etc.). Please note when a category is associated with a particular track or class. Please also note if they make generalizations about race, ethnicity, gender, SES (e.g., parent involvement).*

The summaries and quotes from leaders recorded in response to this prompt provided insight into the Discourses that district leaders invoked to describe various groups of students.

To track mathematical discourses, we drew from the dataset of teacher instruction. The total dataset from 45 teachers over 4 years contained 114 classrooms, with a total of 229 videos (each teacher's classroom was observed twice in either February or March of each year, except for one teacher who had three observations). The lessons often followed a structure of launch, student work time, and then whole-class discussion. As a part of the prior data collection, observed teachers provided student demographic information tied to seating charts and videographers created content logs. The demographic information was used to place students in different subgroups for analysis. For example, if the seating chart identified a student as a "black girl" then we used that information to include that student in the black student subgroup, the girl subgroup, and the black girl subgroup as well. When the information was not provided on the seating chart, we marked the information as missing. The video content logs were used to determine which aspects of the video to code for the present study.

Our analyses focused on whole-class discussions. Our rationale was that while small group work supports student thinking, whole-class discussions are the primary mechanism for advancing a collective mathematical agenda (Speer & Wagner, 2009). In addition, whole-class discussions play an important role in whether or not students are seen publicly as competent (Cohen & Lotan, 2014). Nonetheless, we recognize that small group interactions are another important site of inequity (Ernest et al., 2019). This is a limitation of the current study that can be addressed in future work. Given the complexity and nuanced differences between whole-class and small group work, we decided it was best to limit our focus to whole-class discussions for this paper.

We considered the set of observations associated with a given teacher each year (which were typically subsequent lessons) as a single unit of analysis. For this instructional unit, we determined the overall demographics of the classroom by aggregating the demographics of the students present on each observed day. These instructional units can be considered as a single “classroom,” because even if a teacher was included in the dataset over multiple years of the study, they had different students each year.

<INSERT FIGURE 1>

Analysis

Discourses

Our overall analytic approach for our first research question, *What are district leaders’ discourses about students in relation to mathematics?,* was guided by prior work on racial narratives (Shah, 2017). Although racialized talk does not always invoke racial Discourses, the broader narratives (re)produced by Discourses do tend to emerge through talk about race (and intersectional identities). This unit of analysis aligned with our overarching goal of understanding the impact of Discourses that were present in the district. By understanding the Discourses that leaders were drawing upon broadly, we had a productive tool to make sense of the types of initiatives that were taking place and why, and subsequently, how they might influence and frame teacher practice at the level of mathematics classroom discourse.

Analysis proceeded in three phases. The first phase was to identify and catalog all instances of Discourses about specific groups of students being invoked by participants. Our goal was to identify the content of the Discourses, both considering *who* they refer to, and *what* they were about. In any instance in which a particular group was associated to a particular trait or behavior (e.g., "Asians are good at math") it was coded as representing a Discourse (Shah, 2017). We identified a total of 65 Discourses about groups of students in math.

In the second phase, we attended to the interrelation between Discourses, or so-called "narrative clusters" (Shah, 2017). Rather than considering individual Discourses, we looked at how narratives positioned students vis-à-vis one another. Consider the following cluster,

> She says that the gaps are not surprising and are between middle class kids and kids of poverty. That often corresponds to a racial gap and a gap between native speakers and those whose first language is not English. She believes that "those gaps are all probably based on different things". She believes that "kids coming from poverty and culture has prevented

kids from coming to school at kindergarten or preschool with some of the math knowledge or working with numbers".

This single cluster contains narratives about six groups of students, in three dichotomies: middle-class children vs. those in poverty; students of color vs. white students; and native English speakers vs. English Language Learners. Each of these binaries positions certain groups of students as succeeding at math, while others are not. Notably, students of color, students experiencing poverty, and students learning English are all related, overlapping groups, but they are also distinct. Analyzing Discourses at the unit of clusters allowed us to understand how students were relatively positioned. We identified a total of 33 clusters.

The third and final phase of analysis connected narrative clusters to frames, drawing on prior work (Jackson et al., 2017), rooted in the work of Snow and Benford (1988). We analyzed *explanations* of the relationships between groups (i.e., diagnostic frames) and concrete *actions* in response to those relationships (i.e., prognostic frames). Above, the narrative cluster includes the explanation "kids experiencing poverty have weak math skills." In other narrative clusters, leaders explicitly stated what actions were taken in response to these logics (e.g., providing professional development around ambitious instruction). We identified a total of 29 actions and 17 explanations described by district leaders.

Classroom Discourse: EQUIP Observation Tool

EQUIP is a classroom observation tool for tracking students' talk-based participation in mathematics classrooms (Reinholz & Shah, 2018). EQUIP generates disaggregated data analytics about student participation by combining discourse dimensions (features of participation) with student demographics. The unit of analysis in EQUIP is a *turn*. Each time a new student speaks, it constitutes a new turn. If a single student speaks multiple times without any other students speaking, it would all be considered the same turn. This aggregation allows the EQUIP protocol to adequately capture instances of teacher press.

For this study, in addition to the overall turns by student groups, we focused on three dimensions (solicitation method, teacher question, and explicit evaluation) to understand the nature of those turns. These dimensions were included in EQUIP because of their connections to equitable participation in classroom discourse (Reinholz & Shah, 2018). No tool can capture all meaningful aspects of classroom activity and the EQUIP focuses specifically on verbal participation, given its important role in the learning process (Banes et al., 2019). Nonetheless, we recognize that learners may be engaged in other potentially meaningful forms of participation (i.e. nonverbal), that are not captured (O'Connor et al., 2017).

The first dimension, solicitation method, attends to how the student entered the discussion. Solicitation method describes whether a teacher explicitly *called on* a student, if the student was *not called* on but still spoke, or if a *random* calling method was used. Solicitation method provides insight into the source of inequities in participation, whether they result from explicit teaching moves, or implicit aspects of the classroom environment (i.e., dominant students speaking without being called on).

The second dimension, teacher question type, attends to the nature of the question that was asked that resulted in the student contribution. Teacher question types are *why*, *how*, *what*, *other*, and *N/A* (for no question asked). A *why* question focuses on the reasons behind the math. A *how* question asks for a mathematical process or list of steps. A *what* question requires a student to recall a fact or give an answer. All other questions are coded as *other*. These question types represent a rough hierarchy of quality, with *why* questions generally being more conducive to learning in mathematics classrooms (as they require explanation), and *what* questions typically indicating a lower level of discourse that is only focused on producing answers. Although *what* questions can be used in the service of deeper thinking, if a particular group of students received *only* this type of lower-level question, it would be indicative of inequity.

Finally, the third dimension is explicit evaluation and attends to the way a teacher responded to a student's contribution. Explicit evaluation is coded *yes* or *no*, depending on whether a teacher explicitly judges the quality or correctness of a student's response. The presence of teacher evaluations provide evidence of how authority is distributed between the teacher and students in the classroom (Engle et al., 2014). Combined with *what* type questions, explicit evaluation can signal an Initiation-Response-Evaluation (IRE) pattern of classroom talk (Mehan, 1979). Extensive use of IRE patterns stands in opposition to what would typically be considered academically productive talk (e.g., O'Connor et al., 2017).

While EQUIP captures important aspects of classroom discourse, scaling up traditional discourse analysis necessarily leads to tradeoffs, because discourses focus on situated meanings, resources, and valued practices that may differ across communities and contexts (Moschkovich, 2002). Thus, in our attempt to classify practices across a wide number of classrooms, we necessarily lose some of these more nuanced meanings. For example, we are unable to capture the ways in which students might leverage each other's ideas as they build on them sequentially.

These tradeoffs are necessary, because fine-grained microgenetic discourse analysis that happens at the classroom level is not feasible to implement across a district. What EQUIP offers is flexible, scalable methodology that can be used by researchers and professional developers to understand how race-gender intersections mediate student participation across many classrooms. This methodology has the potential to support both widescale instructional and policy changes, because it allows us to capture broad racial and gender inequities at scale (e.g., the persistent marginalization of Latinx students). Thus, we believe that such methodological tradeoffs can be made strategically in service of equity (Gutierrez, 2002).

Coding and Inter-Rater Reliability.

The coding team consisted of five graduate students, with one *reference coder* who other students were compared to. The rationale behind having a single reference coder is that they were well-versed with the EQUIP protocol from prior studies, and thus provided the "gold standard" that newly trained coders should attempt to replicate. The four *regular coders* were each assigned 24 instructional units (i.e., classrooms), and the remaining classrooms were coded by the reference coder. The reference coder doubled-coded 7 of the 24 classrooms from the other coders, for a total of 56 videos double-coded, or approximately 25% of the dataset. Once all the videos were double coded to acceptable reliability, the coders individually completed coding the rest of their videos. The coding team continued to meet regularly to maintain consistency in approach. To compute interrater reliability, we used Krippendorff's alpha (Hayes & Krippendorff, 2007). We achieved values greater than 0.8 for all raters and all codes (see supplementary materials for exact values, Table A), which is considered good reliability (Carletta, 1996).

Partial transcripts of each lesson were generated to support the coding process. The transcripts included all student talk that was coded, and portions of teacher talk that were relevant to the coding. A total of 3811 student turns were coded using EQUIP. Across this sample, 32 classrooms had full race information, and overall, race or ethnicity information was missing for 17% of students who participated. We dropped 14 classrooms in which race/ethnicity data were missing for over 50% of participating students and analyzed the remaining 100 classrooms. In the classrooms we did analyze, we did not include students with missing race or ethnicity information in the race/ethnicity analysis or the race-gender subgroup analysis. Thus, our analyses included 44 teachers and 3025 coded student turns, in 100 classrooms.

Statistical Analyses.

We conducted a variety of analyses to address the second research question: *How does participation in whole-class mathematical discourse differ by gender and race/ethnicity across the district?* We first focused on participation overall, then by gender and race/ethnicity separately, and finally for race-gender subgroups. To account for differences in the sizes of student demographic groups, we calculated the *average number of turns* by group, by dividing the overall number of turns for each group by the number of students in the group. This allowed us to look at student participation relative to representation in the classroom. For example, if there were 20 turns from Black students in an observation, and 4 Black students in that classroom, the average turns for Black students in that classroom would be 5. That was a number that could then be compared to other demographic groups with different numbers of students (e.g., 20 turns from 20 Latinx students would have an average number of turns of only 1).

With limitations due to space, we choose to engage in deeper analyses of Overall Turns as well as the two dominant codes for the Solicitation Method discourse dimension (Called on Turns and Not Called on Turns) based on visible differences in average terms for different subgroups. Hierarchical linear models were used to account for teachers nested within schools across multiple years. For the hierarchical linear models, we used boy and girl as gender categories, and Asian/Pacific Islander (API), Black, Latinx, White, and Other for racial/ethnic categories (Native and Other were combined into a single category so that the participations patterns of these students would still be accounted for in the model despite small sample sizes). We used robust standard errors to account for heteroskedasticity in the models' unexplained variation. The three-level models followed the general form:

$$TURNS_{ij} = \beta_0 + \beta_1 SUBGROUP(s) + \beta_2 NUM\ Ss + e \qquad (1)$$

$$\beta_q = \pi_{q0} + r_q (q = 0,1,2) \qquad (2)$$

$$\pi_{q0} = \gamma_{q00} + \gamma_{q01} SCHOOL\ \%LEP + \ u_0 (q = 0,1,2) \qquad (3)$$

We included the standardized control variable for the number of students in the subgroup (NUM Ss) because the average number of turns per student was negatively correlated with the number of students: when there are more students in a particular group, it is generally harder to give each of them more turns to participate. We also included the school-level percentage of students identified as limited English proficiency (SCHOOL %LEP) because it was theoretically significant due to our emphasis on classroom discourse, accounted for school level variance, and was significantly related to some of the outcome variables (see equation (3)). We initially included teachers' years of experience, gender, and race/ethnicity (as "White" or "other" due to the large percentage of White teachers) but given a lack of significant correlations between these teacher characteristics and participation outcomes, we removed them from the final models to reduce the number of independent variables in the model.

Overall, the fact that teachers participated in multiple years accounts for a significant proportion of the variance in each of the three outcome variables (see intra-class correlations in Table 1). Further, the inclusion of the school-wide percentage of students identified as having limited English proficiency reduced the proportion of variance explained by school-level clustering when it was initially present.

$$TURNS_{ij} = \beta_0 + \beta_1 BOYS + \beta_2 NUM\ Ss + e \qquad (4)$$

$$TURNS_{ij} = \beta_0 + \beta_1 BLACK + \beta_2 LATINX + \beta_3 API + \beta_4 OTHER + \beta_5 NUM\ Ss + e \quad (5)$$

$$TURNS_{ij} = \beta_0 + \beta_1 OTHER\ GIRLS + \beta_2\ API\ GIRLS + \beta_3\ BLACK\ GIRLS + \beta_4\ WHITE\ GIRLS + \beta_5 OTHER\ BOYS + \beta_6\ API\ BOYS + \beta_7\ BLACK\ BOYS + \beta_8\ LATINX\ BOYS + \beta_9\ WHITE\ BOYS + \beta_{10} NUM\ Ss + e \quad (6)$$

Because of relatively large differences in means by race/ethnicity and gender for classroom average turns of talk, we compared race/ethnicity subgroups (with White students as the comparison group) using one hierarchical linear model (see equation (4)) and gender subgroups (with girls as the comparison group) in a second hierarchical linear model (see equation (5)). We similarly noticed differences in subgroup means for average turns when students were *called on* or *not called on*, so we created hierarchical linear models exploring separate subgroup differences for race/ethnicity and gender.

Finally, we conducted hierarchical linear models with race-gender subgroups (see equation (6)) for each of those three outcomes: 1) average turns of talk, 2) average turns when students were called on, and 3) average turns when student were not called on. We used Latinx girls as our comparison group because of the high likelihood of their being lower than other subgroups, based on prior models.

Results

Discourses

Discourses and Discourse Clusters

Consistent with broader literature on urban schools, a variety of Discourses were invoked to make sense student performance. Broadly speaking, the leaders framed student performance in terms of achievement gaps. They associated “higher performance” with “middle-class white students”, in contrast to “struggling students” and “students experiencing poverty.” These latter two categories were associated with students of color generally, and Black students in specific. Black students were the most frequently referenced group in Discourses (18/65, or 27.7% of mentions). The next most frequently mentioned group was white students (11/65, or 17%). When white students were discussed, it was only in narrative clusters that positioned them relative to students of color. Other racial groups were mentioned infrequently, with five mentions of Latinx students and four mentions of Native American students, and these mentions were coincident with the description of Black students, as part of a collective *othering* of students of color in achievement gap rhetoric.

Additionally, “students experiencing poverty” (9/65, or 14%) were mentioned frequently. In almost all these mentions, students experiencing poverty and students of color or Black students were described as overlapping groups. The only other group that was mentioned frequently was ELLs (6/65, or 9%). ELLs were described separately from students of color, even though in fact the ELLs in the district were primarily students of color. Notably absent was any mention of gender narratives in mathematics.

Explanations (Diagnostic Frames)

Of the explanations we coded, 16/17 (94%) of them specifically related to Black students or students of color. These could broadly be separated into two categories: student deficits (5/16, 31%) and systemic issues (11/16, 69%). Explanations of student deficits included that they "had a math phobia," that they "don't have the vocabulary to handle the [district-adopted] curriculum," or missing prerequisite "math knowledge." What is more notable, is that the leaders typically framed the gaps in terms of institutional barriers and teacher expectations. For example, they described, "kids are tracked at younger grades," "primarily white teachers lack expectations that kids of color can do well," and a need to "remove the gateways that adults put up." It's worth noting that only a single explanation focused on ELLs, in which a leader described the families of East African immigrant populations as "more coherent" than other ELL populations.

Actions (Prognostic Frames)

Leaders also described actions that the districts were taking to address inequities. Most of these were targeted towards supporting Black students in specific, and students of color in general (19/29, 65%). For instance, an initiative focused on a particular neighborhood of the district was a "comprehensive effort" that "coordinates with social services," including "housing, schools with afterschool and Saturday programming, [and] health care." This holistic effort aimed to address a variety of different social circumstances of Black students living in poverty in the district. The district was also collaborating with the National Urban Alliance (NUA), "which emphasizes the use of culturally relevant curriculum" and "sends consultants" to "work with teachers" for a few days about four times a year. The NUA was organized around Positive Behavior Intervention Systems and the use of data to provide support.

To address teacher expectations about Black students, the district was collaborating with Courageous Conversations. The Assistant Director for Professional Development elaborated:

> *Courageous Conversations has a package of ways to think about one's own belief system and racism…[we] had 6 sessions with them… [to] embed the Courageous Conversations in [professional development] with teachers/principals over the course of the year…. In the curriculum guide, they are going to make sure that they provide suggestions for how*

> *teachers can set up the context of the curriculum so that kids can access it (prior experience, vocabulary).*

Another district leader elaborated how they had included a session on culturally responsive pedagogy and focused on the ways in which they could integrate it with accountable talk moves (i.e., ambitious instruction).

Other efforts included the use of Sheltered Instruction to support ELLs, and the adoption of the Saxon textbook for students in special education. In all, fewer actions discussed focused on ELLs (7/29), and students in special education (3/29).

Summary of Discourse Analysis

The analysis of broad Discourses in the district revealed that the achievement gap was the primary lens through which leaders viewed student performance. While they noted gaps for ELLs, the primary focus was on racial disparities, specifically between Black and white students. Despite the problematic nature of achievement gap Discourses writ large (Gutiérrez, 2008), leaders primarily framed the gaps in terms of institutional barriers and teacher expectations, not student deficits. This framing has the potential to lead to "justice in the gaps" as it addresses disparities in opportunities (Flores & Gunzenhauser, 2021). The result was a variety of targeted interventions aimed to improve cultural competence, teaching quality, and provide holistic support for Black students. We now turn our attention to classroom mathematical discourse.

Classroom mathematical discourse

Our second area of analysis focused on students' mathematical discourse. A summary of discourse dimensions is given in Table 2. First, we summarize the results overall. Most turns (81%) occurred when a teacher *called on* students, but this still left a considerable portion of turns to students sharing ideas when *not called on* (18.7%). Teachers rarely, if ever, used *random* calling methods. Most of the students' contributions were in response to teachers' questions (94.6%). And teachers tended to ask low-level *what* type of questions (64.5%) (see Table 2). We were interested in whether there were differences in who gets asked "why" questions, but there were so few why questions in general, it was not possible detect significant differences in who gets asked why questions with this data set. Although teachers engaged students in low-level discourse, most of the time the teachers *did not evaluate* the responses given by students (78.6%), indicating that students had some opportunities to act as authorities in the math classroom. Perhaps this lack of evaluation contributed to prior results showing a comparatively high level of academically productive talk in the district (Boston & Wilhelm, 2017). In Table 2, we provide informtion about the frequency of each code within each dimension, disaggregated by sub-group.

<INSERT TABLE 2>

Table 3 shows that participation appears to vary across gender and racial groups. For instance, boys had 2.1 turns on average, compared to 1.72 for girls. Average turns differed by race, with Black (1.92) and White (1.80) having the highest numbers, and API (1.17) and Latinx (1.14) having less participation. Our analyses showed that three racial groups – Latinx, API, and Native American – often had no observed participation in whole class discussions across the multiple lessons we coded (~25%, ~50%, and ~75% of classrooms in which they were present, respectively). The number of classrooms where students identified within each racial/ethnic category were present are listed at the top of Table 3. For example, there were 55 classrooms (of 100) with students identified as Asian/Pacific Islander students presents. And, in 50% of those classrooms those students did not participate. Unfortunately, given that only 14 of the 100 classrooms contained students identified as Native American and only 3 classrooms had Native American students who participated, we could not explore these differences statistically. The distributions of participation are given in Figure 2.

<INSERT TABLE 3>

<INSERT FIGURE 2>

Subgroup models

The first set of models focused on exploring whether visual differences between group means were statistically significant, controlling for the number of students in the subgroup, school level percentage of students identified as having limited English proficiency and teacher-level and school-level dependencies. We found that differences in the average number of turns for boys and girls were statistically significant ($p<.001$). When looking at the type of Solicitation Method, we found that boys had significantly more turns when *not called on* ($p<0.001$), and the difference was marginally significant when boys were *called on* ($p=0.05$). Comparing average turns for subgroups of different race or ethnicity, we found that the average number of turns for Latinx students was significantly lower than the average number of turns for White and Black students ($p<.001$ and $p<.05$, respectively). The pattern was similar for Asian/Pacific Islander students with significantly fewer turns of talk than White and Black students ($p<.05$ for both groups). The difference between average number of turns for White and Black students was not statistically significant. These differences by race/ethnicity were mirrored for turns of talk when *called on*, with Asian/Pacific Islander and Latinx students having fewer average turns of talk. With respect to when *not called on*, the pattern was different: Black students had significantly more turns of talk than the other racial/ethnic groups ($p<.05$), and there were no significant differences between the other racial/ethnic groups.

Race-Gender Subgroup Models

For the race-gender subgroup models, we first examined the difference in *average turns of talk* race-gender subgroups. Our findings again revealed significant differences between boys and girls. All the means for race-gender subgroups with boys were significantly greater than the mean for Latinx girls, and the means for all the other subgroups with girls were not statistically significantly different from the mean for Latinx girls (see Table 4). The average number of turns for Latinx girls (the comparison group, so reported as the constant in the results table) was 1.81. All boys had significantly more turns to talk, with boys having around 0.40 more turns, so approximately 2.21 turns overall ($p<.05$).

<INSERT TABLE 4>

The subgroup models for turns when students were *called on* and *not called on* revealed distinctive patterns of differences between race-gender subgroups. First, recall that in the sample classrooms, students were called on much more often than not. We found that Latinx girls had on average 0.99 *called on* turns. This was statistically at the same level of participation as Latinx boys and Asian/Pacific Islander girls and boys. From there, Black and White girls had about 0.60 more turns (b=0.65, b=0.59, respectively, $p<.05$), or 1.59 average *called on* turns. Finally, Black and White boys had about 0.96 more turns (b=0.95, b=0.98, respectively, $p<.001$), or 1.95 average *called on* turns.

In the race-gender models for when students were *not called on*, we found different results. We found that Latinx girls had on average 0.14 turns when *not called on*. This mean was statistically significantly different from the means for three groups. It was significantly smaller than the means for Black boys and Black girls (b=0.47 and b=0.15, respectively, $p<.001$ and $p<.05$). Therefore, Black boys had approximately 0.61 turns when not called on and Black boys had approximately 0.29 turns when not called on. Based on a parallel model with Black boys as the comparison group, the difference between White and Black boys' turns when not called on was statistically significant ($p<.05$). The mean for boys of other races or ethnicities was significantly lower than the mean for Latinx girls (b = -0.09, $p<.05$), meaning that boys of other races or ethnicities had a mean of 0.05 turns when *not called on*. This result is difficult to interpret confidently because of the variation within the "other" category and the small magnitude of the result. All the other intersectional groups were statistically equivalent to the average turns for Latinx girls when *not called on*.

Discussion

This study provides a picture of student participation by race and gender across 100 middle-school classrooms from a racially diverse urban school district. We intentionally selected an instructionally advanced district with external support and a commitment to ambitious and equitable instruction (Boston & Wilhelm, 2017). Nonetheless, we found that patterns of participation in mathematical discourse were not fairly distributed across gender and racial groups. Our findings of districtwide inequities under these relatively favorable circumstances underscore the ubiquity of racial and gender inequities in middle school mathematics classrooms in the USA. We use a poststructural lens that attends to race-gender Discourses to interpret these results. While we cannot establish that race-gender Discourses *caused* the particular discourse patterns we observed, we do hypothesize that the two are closely related.

Overall, we found boys participated more than girls. Further, we found a statistically significant difference in *not called on* talk—and marginally significant difference in *called on* talk. These patterns are consistent with masculine Discourses in mathematics broadly, which position men as louder, more competitive, and eager to participate publicly (e.g., Leyva, 2017; Lubienski & Ganley, 2017). While we suspect that teachers lacked intentional strategies to elicit comparable participation from girls, the lower levels of *not called on* talk from girls may also have been driven by pressure to engage in compliant "good-girl" behavior (Lubienski & Ganley, 2017). Notably, gender Discourses were *never* invoked by district leaders, and gender was not seen as a significant area of inequity. We find this surprising, given the significant disparities in participation between boys and girls across racial groups.

When we disaggregated by race, we found that Black and White boys had significantly more turns of participation than all Latinx students, both when *called on* and *not called on*. This finding is particularly important, because while Black and Latinx students are often treated monolithically in research work (e.g., under the umbrella of "underrepresented" or "minoritized" students), these groups have unique experiences. Given the observational nature of this quantitative study, we cannot speak directly to students' subjective experiences, but nonetheless we offer some possible explanations for these results.

Our finding that Black students received a proportional share of participation opportunities was notable, especially in light of ongoing anti-Black racism in our society writ large and in mathematics education specifically (Martin et al., 2017). Moreover, Black boys had significantly more *not called on* turns than White boys, the next highest group. This finding also highlights that Black boys and Black girls had different experiences. This is consistent with race-gender Discourses that tend to invisibilize Black girls in comparison to Black boys (Gholson, 2016).

We speculate that the relatively higher levels of participation of Black students was due to the increased attention Black students received in school- and district-based initiatives. District leaders appeared to have a laser focus on reducing perceived gaps between Black and White students, and this manifested in an array of interventions focused both on supporting students holistically (e.g., the neighborhood success initiative) and improving teacher's cultural competence (Webster, 2016). In this case, the leaders framed the gaps in terms of structural barriers and institutional policies appeared to provide effective grist for improvement (Flores & Gunzenhauser, 2021). To be clear, though, we know little of the subjective experiences of Black students in the district, and do not take these data as a sign that inequities have been eliminated. Rather, these findings may provide some preliminary evidence of *improved* equity, warranting follow-up research.

In contrast, we found that Latinx students of all genders had the lowest levels of participation, drawing attention to intersectional oppression. This finding is consistent with lack of attention to Latinx students in Discourses within the district, and broader Discourses that invisibilize Latinx students. We suspect that stereotypes of Latinx students not caring about school and negative perceptions of their language ability may work together to render them academically invisible in mathematical discussions. These stereotypes intersect with gender stereotypes about Latinas as "well-behaved" or "fragile," providing them even less attention from teachers (Shah, Herbel-Eisenmann, et al., 2020). Across all models, Latinx girls had the lowest levels of participation. When all forms of participation were considered, Latinx boys participated more than Latinx girls, but when *called on* and *not called on* turns were considered separately, no statistical differences could be detected between Latinx girls and boys.

It was also notable that API students had relatively low levels of participation across the sample, at levels comparable to Latinx populations. Our finding that API students had a relatively modest level of participation stands in contradiction to conventional Discourses and assumptions about Asian student success and highlights the importance of looking more carefully at different subpopulations. Here, the lower levels of participation might be related to the sizable Hmong immigrant community within the sample, which differs in important ways for other Asian populations (Pang et al., 2011). Thus, when research does not disaggregate Asian subpopulations, it can further perpetuate the marginalization and invisibilization of subpopulations that are not often considered in mathematics education research.

Our study has a few key limitations. First, we were limited to the gender and racial/ethnic categories within the existing dataset therefore we could only analyze two gender categories and were limited in the racial/ethnic categorizations (e.g., we could not fully disaggregate Asian or Black subpopulations). Second, our sample and space constraints limited the analyses we could perform. For example, while we found evidence of low levels of participation from Native students, we were not able to explore these differences statistically. Also, we chose to provide a broad descriptive picture of discussion patterns in classrooms, and then statistically analyze Overall Turns and Solicitation Method further to understand whether perceived differences based on the descriptive statistics were statistically significant. Third, although we considered race-gender subgroups, we still recognize the limitation of these rigid categories, which can obscure important within-population variation. Fourth, the teachers in the sample were primarily White. Although there is some evidence that teachers of color also hold biases in favor of White students (Copur-Gencturk et al., 2019), we caution about overgeneralizing to different populations of teachers. Nonetheless, we also note that teacher race was not statistically significant in our hierarchical linear models. Fifth, our analysis of district Discourses and classroom discourse leaves out an important middle ground of school or classroom Discourses which is beyond the scope of this paper but is likely influential, and would account for another source of variation in the sample (Jackson, 2009).

Our findings have practical significance for teachers and teacher educators in urban mathematics education settings. We provide empirical evidence of widespread inequities in mathematics discussions across race-gender subgroups. Because teachers have a large degree of control over how they structure discussions and how they solicit participation, we argue that individual classrooms can be an important site for change in urban mathematics education.

Simultaneously, classrooms do not operate in vacuum, and systemic policy shifts are needed to empower teachers to succeed in this work.

Our study suggests that racial equity initiatives with a focus on Black students may have had a positive impact on the experiences of Black learners in this district. Nonetheless, these benefits were unequal for Black boys and Black girls. This may have been driven in part due to the lack of attention at the district level to gender as a source of marginalization in mathematics. In this way, our study highlights the need for an intersectional approach that attends to interlocking oppressive Discourses; when racial Discourses are considered in isolation (e.g., without attention to gender as well), it may only further reinscribe different types of inequity. Moreover, our findings showed that not all racially minoritized groups benefited equally from the district efforts, which highlights the need for more targeted and nuanced forms of professional learning that can address the subtle and pervasive forms of racialization that are present in mathematics classrooms.

We suspect that providing teachers and district leaders with the same types of data that were used in this study (i.e., disaggregated classroom participation data), could provide productive grist for professional learning and policy change that attends to interlocking systems of oppression. This approach has shown preliminary promise at smaller scales (Reinholz et al., 2019; Shah, Christensen, et al., 2020), and a future step will be to apply such work at the scale of schools and school districts.

References

Annamma, S. A., Anyon, Y., Joseph, N. M., Farrar, J., Greer, E., Downing, B., & Simmons, J. (2019). Black Girls and School Discipline: The Complexities of Being Overrepresented and Understudied. *Urban Education*, *54*(2), 211–242. https://doi.org/10.1177/0042085916646610

Banes, L. C., Restani, R. M., Ambrose, R. C., Martin, H. A., & Bayley, R. (2019). Relating Performance on Written Assessments to Features of Mathematics Discussion. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-019-10029-w

Battey, D., & Leyva, L. A. (2016). A framework for understanding whiteness in mathematics education. *Journal of Urban Mathematics Education*, *9*(2).

Berry, R. Q., Ellis, M., & Hughes, S. (2014). Examining a history of failed reforms and recent stories of success: Mathematics education and Black learners of mathematics in the United States. *Race Ethnicity and Education*, *17*(4), 540–568. https://doi.org/10.1080/13613324.2013.818534

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

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, *22*(2), 249–254.

Carter Andrews, D. J., Brown, T., Castro, E., & Id-Deen, E. (2019). The Impossibility of Being "Perfect and White": Black Girls' Racialized and Gendered Schooling Experiences. *American Educational Research Journal*, *56*(6), 2531–2572. https://doi.org/10.3102/0002831219849392

Cobb, P., Jackson, K., Henrick, E., & Smith, T. M. (2020). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Harvard Education Press.

Cohen, E. G., & Lotan, R. A. (2014). *Designing Groupwork: Strategies for the Heterogeneous Classroom Third Edition*. Teachers College Press.

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*, 0013189X19890577. https://doi.org/10.3102/0013189X19890577

Crenshaw, K. (1990). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), 1241–1300.

Engle, R. A., Langer-Osuna, J. M., & Royston, M. M. de. (2014). Toward a Model of Influence in Persuasive Discussions: Negotiating Quality, Authority, Privilege, and Access Within a Student-Led Argument. *Journal of the Learning Sciences*, *23*(2), 245–268. https://doi.org/10.1080/10508406.2014.883979

Ernest, J. B., Reinholz, D. L., & Shah, N. (2019). Hidden competence: Women's mathematical participation in public and private classroom spaces. *Educational*

Studies in Mathematics, *102*(2), 153–172. https://doi.org/10.1007/s10649-019-09910-w

Faulkner, V. N., Stiff, L. V., Marshall, P. L., Nietfeld, J., & Crossland, C. L. (2014). Race and Teacher Evaluations as Predictors of Algebra Placement. *Journal for Research in Mathematics Education*, *45*(3), 288–311. https://doi.org/10.5951/jresematheduc.45.3.0288

Flores, O. J., & Gunzenhauser, M. G. (2021). Justice in the Gaps: School Leader Dispositions and the Use of Data to Address the Opportunity Gap. *Urban Education*, *56*(2), 261–288. https://doi.org/10.1177/0042085918801431

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Pantheon Books.

Gholson, M. L. (2016). Clean Corners and Algebra: A Critical Examination of the Constructed Invisibility of Black Girls and Women in Mathematics. *The Journal of Negro Education*, *85*(3), 290–301. JSTOR. https://doi.org/10.7709/jnegroeducation.85.3.0290

Goldin, S., Khasnabis, D., O'Connor, C., & Hearn, K. (2021). Tangling With Race and Racism in Teacher Education: Designs for Counterstory-Based Parent Teacher Conferences. *Urban Education*, *56*(9), 1399–1428. https://doi.org/10.1177/0042085919894042

Gutierrez, R. (2002). Enabling the Practice of Mathematics Teachers in Context: Toward a New Equity Research Agenda. *Mathematical Thinking and Learning*, *4*(2 & 3), 145–187.

Gutiérrez, R. (2008). A "Gap-Gazing" Fetish in Mathematics Education? Problematizing Research on the Achievement Gap. *Journal for Research in Mathematics Education*, *39*(4), 357–364.

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2015). Student participation in elementary mathematics classrooms: The missing link between teacher practices and student achievement? *Educational Studies in Mathematics*, *90*(3), 341–356.

Jackson, K. (2009). The social construction of youth and mathematics: The case of a fifth grade classroom. In D. B. Martin (Ed.), *Mathematics teaching, learning, and liberation in the lives of Black children* (pp. 175–199). Routledge.

Jackson, K., Gibbons, L., & Sharpe, C. J. (2017). Teachers' Views of Students' Mathematical Capabilities: Challenges and Possibilities for Ambitious Reform. *Teachers College Record*, *119*(7), 1–43. https://doi.org/10.1177/016146811711900708

Jett, C. C. (2019). Mathematical Persistence Among Four African American Male Graduate Students: A Critical Race Analysis of Their Experiences. *Journal for Research in Mathematics Education*, *50*(3), 311–340. JSTOR.

Lappan, G., Fey, J., Fitzgerald, W. M., Friel, S. N., & Phillips, E. D. (2003). *Connected mathematics*. Pearson Prentice Hall.

Larnell, G. V., & Martin, D. B. (2021). Urban Mathematics Education as a Political and Personal Project. In *Handbook of Urban Education* (2nd ed.). Routledge.

Leath, S., Mathews, C., Harrison, A., & Chavous, T. (2019). Racial Identity, Racial Discrimination, and Classroom Engagement Outcomes Among Black Girls and Boys in Predominantly Black and Predominantly White School Districts. *American Educational Research Journal*, *56*(4), 1318–1352. https://doi.org/10.3102/0002831218816955

Leyva, L. A. (2017). Unpacking the male superiority myth and masculinization of mathematics at the intersections: A review of research on gender in mathematics education. *Journal for Research in Mathematics Education*, *48*(4), 397–433. https://doi.org/10.5951/jresematheduc.48.4.0397

Lubienski, S. T., & Ganley, C. M. (2017). Research on gender and mathematics. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 649–666). National Council of Teachers of Mathematics.

Martin, D. B. (2012). Learning Mathematics while Black. *Educational Foundations*, *26*, 47–66.

Martin, D. B., Rousseau-Anderson, C., & Shah, N. (2017). Race and mathematics education. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 607–636). National Council of Teachers of Mathematics.

Mehan, H. (1979). "What Time Is It, Denise?": Asking Known Information Questions in Classroom Discourse. *Theory into Practice*, *18*(4), 285–294.

Milner, H. R. (2012). But What is Urban Education? *Urban Education*, *47*(3), 556–561. https://doi.org/10.1177/0042085912447516

Moschkovich, J. (2002). A Situated and Sociocultural Perspective on Bilingual Mathematics Learners. *Mathematical Thinking and Learning*, *4*(2–3), 189–212. https://doi.org/10.1207/S15327833MTL04023_5

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*(1), 187–240.

Oakes, J. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science.* Rand Corp. http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED329615

O'Connor, C., Michaels, S., Chapin, S., & Harbaugh, A. G. (2017). The silent and the vocal: Participation and learning in whole-class discussion. *Learning and Instruction*, *48*, 5–13. https://doi.org/10.1016/j.learninstruc.2016.11.003

Pang, V. O., Han, P. P., & Pang, J. M. (2011). Asian American and Pacific Islander Students: Equity and the Achievement Gap. *Educational Researcher*, *40*(8), 378–389. https://doi.org/10.3102/0013189X11424222

Reinholz, D. L. (2022). Interrogating Innate Intelligence Racial Narratives: Students' Construction of Counter-Stories within the History of Mathematics. *International Journal of Research in Undergraduate Mathematics Education*, *8*(1), 36–63. https://doi.org/10.1007/s40753-021-00145-w

Reinholz, D. L., Johnson, E., Andrews-Larson, C., Stone-Johnstone, A., Smith, J., Mullins, B., Fortune, N., Keene, K., & Shah, N. (2022). When Active Learning Is Inequitable: Women's Participation Predicts Gender Inequities in Mathematical

Performance. *Journal for Research in Mathematics Education*, *53*(3), 204–226. https://doi.org/10.5951/jresematheduc-2020-0143

Reinholz, D. L., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education*, *49*(2), 140–177.

Reinholz, D. L., Stone-Johnstone, A., & Shah, N. (2019). Walking the walk: Using classroom analytics to support instructors to address implicit bias in teaching. *International Journal for Academic Development*, *0*(0), 1–14. https://doi.org/10.1080/1360144X.2019.1692211

Sengupta-Irving, T. (2021). Positioning and Positioned Apart: Mathematics Learning as Becoming Undesirable. *Anthropology & Education Quarterly*, *52*(2), 187–208. https://doi.org/10.1111/aeq.12378

Shah, N. (2017). Race, Ideology, and Academic Ability: A Relational Analysis of Racial Narratives in Mathematics. *Teachers College Record*, *119*(7), 1–42.

Shah, N., Christensen, J. A., Ortiz, N. A., Nguyen, A., Byun, S., Stroupe, D., & Reinholz, D. L. (2020). Racial hierarchy and masculine space: Participatory in/equity in computational physics classrooms. *Computer Science Education*, *0*(0), 1–25. https://doi.org/10.1080/08993408.2020.1805285

Shah, N., Herbel-Eisenmann, B., & Reinholz, D. L. (2020). Why Mrs. Stone never calls on Debra: A case of race-gender ideology in practice. In M. Gresalfi & I. S. Horn (Eds.), *The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020, Volume 1* (pp. 1974–1981). International Society of the Learning Sciences.

Snow, D. A., & Benford, R. D. (1988). Ideology, frame resonance, and participant mobilization. *International Social Movement Research*, *1*(1), 197–217.

Speer, N. M., & Wagner, J. F. (2009). Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *Journal for Research in Mathematics Education*, *40*(5), 530–562. https://doi.org/10.5951/jresematheduc.40.5.0530

Stinson, D. W. (2013). Negotiating the “White Male Math Myth”: African American Male Students and Success in School Mathematics. *Journal for Research in Mathematics Education*, *44*(1), 69–99. https://doi.org/10.5951/jresematheduc.44.1.0069

Taylor, J., Kyere, E., & King, È. (2021). A Gardening Metaphor: A Framework for Closing Racial Achievement Gaps in American Public Education System. *Urban Education*, *56*(9), 1547–1575. https://doi.org/10.1177/0042085918770721

Voigt, M., & Reinholz, D. L. (2020). *Calculating Queer Acceptance and Visibility: A Literature Synthesis on Queer Identity in Mathematics*. Open Science Framework. https://doi.org/10.31219/osf.io/pumqe

Wai, J. (2012). *Why Is It Socially Acceptable To Be Bad At Math?* Psychology Today. http://www.psychologytoday.com/blog/finding-the-next-einstein/201203/why-is-it-socially-acceptable-be-bad-math

Webster, M. (2016). *A decomposition of the practices of high quality professional development facilitation for teachers*. McGill University Libraries.

Wilkes, C. E., & Ball, D. L. (2020). *Positioning, what do we know? An investigation of how Black learners are positioned in mathematics classrooms and the role of* *intersectionality as an analytical tool for understanding why* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/ycv34

Table 1

Intra-class Correlation Coefficients for the Null and Conditional (on School %LEP) Models

Level	Average Turns	Average Turns (With School %LEP)	Called On Turns	Called on Turns (with %LEP)	Not Called On Turns	Not Called On (with %LEP)
School	0.068	0.052	0	0	0.014	0.002
Teacher within School	0.324	0.315	0.412	0.406	0.357	0.354

Table 2

Intersectional data by race and gender

		Overall	African American		Latinx		White	
			Boys	**Girls**	**Boys**	**Girls**	**Boys**	**Girls**
	All Contributions	3469	726	587	261	214	599	491
Solicitation Method	**Called On**	2861 (82.5%)	554 (76.3%)	481 (81.9%)	219 (83.9%)	185 (86.4%)	490 (81.8%)	457 (93.1%)
	Not Called On	568 (16.4%)	168 (23.1%)	102 (17.4%)	42 (16.1%)	29 (13.6%)	108 (18.0%)	34 (6.9%)
	Random	19 (0.5%)	4 (0.6%)	4 (0.7%)	0	0	1 (0.2%)	0
Teacher Question	**Why**	434 (12.5%)	99 (13.6%)	66 (11.2%)	31 (11.9%)	31 (14.5%)	74 (12.4%)	55 (11.2%)
	How	303 (8.7%)	67 (9.2%)	54 (9.2%)	21 (8.1%)	15 (7.0%)	47 (7.8%)	50 (10.2%)
	What	2238 (64.5%)	445 (61.3%)	388 (66.1%)	164 (62.8%)	140 (65.4%)	375 (62.6%)	317 (64.6%)
	Other	307 (8.8%)	68 (9.4%)	50 (8.5%)	28 (10.7%)	14 (6.5%)	56 (9.3%)	56 (11.4%)
	NA	187 (5.4%)	47 (6.5%)	29 (4.9%)	17 (6.5%)	14 (6.5%)	47 (7.9%)	13 (2.6%)
Explicit Evaluation	**Yes**	742 (21.4%)	539 (74.2%)	451 (76.8%)	203 (77.8%)	171 (79.9%)	489 (81.6%)	430 (87.6%)
	No	2727 (78.6%)	187 (25.8%)	136 (23.2%)	58 (22.2%)	43 (20.1%)	110 (18.4%)	61 (12.4%)

*Random was dropped from the statistical analyses given its relative infrequency in the dataset.

Table 3

Average number of turns (and standard deviation) of discourse dimensions by demographic group (N = number of classrooms for which each group was present).

		Overall	Boy N=100	Girl N=100	API N=55	Black N=98	Latinx N=77	White N=87
Number of students			9.64 (3.74)	10.19 (3.94)	1.16 (1.58)	4.05 (4.31)	7.37 (4.39)	6.86 (7.42)
Percentage of classrooms where this group participated			100%	100%	52.7%	72.7%	94.9%	86.2%
	Average Turns	1.87 (1.16)	2.10 (1.37)	1.72 (1.26)	1.17 (1.96)	1.92 (1.78)	1.14 (1.41)	1.80 (1.70)
Solicitation Method	**Called On**	1.52 (1.04)	1.64 (1.16)	1.46 (1.23)	0.88 (1.30)	1.56 (1.65)	0.97 (1.23)	1.58 (1.57)
	Not Called On	0.35 (0.44)	0.45 (0.63)	0.25 (0.38)	0.27 (0.97)	0.36 (0.51)	0.17 (0.40)	0.22 (0.38)
	Random	0.01 (0.07)	0.01 (0.07)	0.01 (0.07)	0.02 (0.16)	0.01 (0.05)	0 --	0.00 (0.01)
Teacher Question	**Why**	0.22 (0,25)	0.25 (0.31)	0.20 (0.25)	0.18 (0.42)	0.25 (0.50)	0.11 (0.19)	0.18 (0.34)
	How	0.16 (0.19)	0.19 (0.27)	0.14 (0.19)	0.09 (0.25)	0.16 (0.21)	0.09 (0.22)	0.16 (0.28)
	What	1.23 (0.96)	1.35 (1.07)	1.17 (1.13)	0.68 (1.06)	1.25 (1.30)	0.79 (1.24)	1.21 (1.37)
	Other	0.16 (0.24)	0.19 (0.28)	0.13 (0.24)	0.14 (.47)	0.18 (0.36)	0.09 (0.21)	0.16 (0.34)
	NA	0.10 (0.17)	0.12 (0.24)	0.07 (0.14)	0.07 (0.31)	0.09 (0.23)	0.05 (0.19)	0.09 (0.21)
Explicit Evaluation	**Yes**	0.40 (0.43)	0.49 (0.55)	0.33 (0.39)	0.34 (0.76)	0.40 (0.49)	0.25 (0.54)	0.34 (0.54)
	No	1.47 (1.01)	1.61 (1.14)	1.38 (1.15)	0.83 (1.35)	1.52 (1.59)	0.89 (1.05)	1.46 (1.50)

*A breakdown of dimensions is not shown for Native students. Of the 14 classrooms in which Native students were present (1.7 students on average), we only coded participation for Native students in 3 classrooms.

Table 4

Hierarchical Linear Models for Race-Gender Subgroups.

	Model 1: All Turns		Model 2: Called on Turns		Model 3: Not Called On Turns	
	b	SE	b	SE	b	SE
Other Girls	-0.06	0.03	-0.62**	0.13	0.01	0.06
A/PI Girls	-0.04	0.03	-0.25	0.18	0.001	0.07
Black Girls	0.06*	0.02	0.65*	0.21	0.15*	0.06
White Girls	0.05	0.04	0.59*	0.29	0.04	0.07
Other Boys	0.32*	0.10	-0.66*	0.32	-0.09*	0.04
A/PI Boys	0.33**	0.09	0.26	0.31	0.26	0.13
Black Boys	0.43**	0.08	0.95**	0.22	0.48**	0.13
Latinx Boys	0.38**	0.09	0.05	0.16	0.09	0.06
White Boys	0.42**	0.11	0.98**	0.24	0.26	0.13
Num Ss	-0.09*	0.03	-0.17*	0.07	-0.02	0.03
School %LEP	0.15	0.19	0.21	0.12	0.06	0.04
Constant	1.81**	0.18	0.99**	0.13	0.14**	0.04
Variance Components						
School	0.037		0		0	
Teacher within School	0.522		0.235		0.090	

Figure Captions

Figure 1. *Racial breakdown of student populations across the ten schools observed in District A*.

Figure 2. *Distributions of average turns by group.*

Supplementary Materials.

Table A. Krippendorf's Alpha for different raters and dimensions

	Solicitation Method	Teacher Question	Explicit Evaluation
Rater 1 (N=279)	0.889	0.912	0.861
Rater 2 (N=212)	0.930	0.815	0.818
Rater 3 (N=352)	0.876	0.819	0.839
Rater 4 (N=183)	0.893	0.817	0.838

Table B. Student Demographics of School Sites.

School	% API M(SD)	% Black M(SD)	% Latinx M(SD)	% White M(SD)	% Native American M(SD)	% English Language Learner M(SD)	Number of Students*	Title I School
1	15 (4.24)	47.6 (6.47)	11.8 (11.98)	22.8 (3.42)	2.8 (0.45)	13.4 (13.1)	470	Yes
2	8.4 (1.52)	21 (1.22)	3.8 (0.83)	65.2 (2.77)	0.8 (0.84)	6.2 (5.67)	600	No
3	7.2 (0.84)	21.6 (2.88)	6 (1.41)	63 (3.16)	1.8 (0.45)	1.4 (1.67)	490	No
4	3.6 (0.89)	31.4 (2.30)	50.2 (3.35)	11.4 (1.14)	3.8 (0.45)	28.6 (26.15)	410	Yes
5	8.8 (2.59)	56.8 (1.30)	30.2 (0.83)	22.0 (1.09)	2.2 (1.79)	16.8 (15.48)	470	Yes
6	7 (0.71)	43.8 (1.30)	18.4 (0.55)	26.2 (1.09)	4 (0)	9.6 (8.85)	480	Yes
7	21.2 (2.28)	63.4 (5.03)	4.0 (1.22)	10 (2.82)	2.2 (0.84)	9.2 (8.53)	360	Yes
8	3.2 (0.45)	35.8 (1.92)	37.0 (3.94)	20.6 (4.51)	3.2 (0.84)	19.2 (17.75)	730	Yes
9	2.0 (0.71)	50.2 (11.03)	10.0 (3.32)	31.2 (5.89)	6.6 (1.52)	16.0 (15.02)	440	Yes
10	3.4 (1.14)	59.8 (3.11)	6.6 (1.14)	12.8 (4.02)	15.2 (3.11)	17.0 (15.52)	500	Yes

- Rounded to the nearest multiple of 10