Journal for Research in Mathematics Education

The Journal for Research in Mathematics Education is an official journal of the National Council of Teachers of Mathematics (NCTM). *JRME* is the premier research journal in mathematics education and is devoted to the interests of teachers and researchers at all levels—preschool through college.

ARTICLE TITLE:

When Active Learning Is Inequitable: Women's Participation Predicts Gender Inequities in Mathematical Performance

AUTHOR NAMES:

Reinholz, Daniel; Johnson, Estrella; Andrews-Larson, Christine;

DIGITAL OBJECT IDENTIFIER: 10.5951/jresematheduc-2020-0143

VOLUME: 53

ISSUE NUMBER: 3

Mission Statement

The National Council of Teachers of Mathematics advocates for high-quality mathematics teaching and learning for each and every student.

CONTACT: jrme@nctm.org

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS


JULY 2019 NUMBER 4 VOLUME 50

Journal for Research in Mathematics Education

EDITORIAL

Choosing and Justifying Robust Methods for Educational Research — 342
Jinfa Cai, Anne Morris, Charles Hohensee, Stephen Hwang, Victoria Robison, Michelle Cirillo, Steven L. Kramer, and James Hiebert

RESEARCH COMMENTARY

Unpacking the Links Between Equitable Teaching Practices and Standards for Mathematical Practice: Equity for Whom and Under What Conditions? — 349
Filiberto Barajas-López and Gregory V. Larnell

RESEARCH REPORTS

Investigating Teaching in Conceptually Oriented Mathematics Classrooms Characterized by African American Student Success — 362
Jonee Wilson, Mahtab Nazemi, Kara Jackson, and Anne Garrison Wilhelm

Anti-Deficit Narratives: Engaging the Politics of Research on Mathematical Sense Making — 401
Aditya P. Adiredja

Lived Moments of Shift in Prospective Elementary Teachers' Mathematical Learning — 436
Keri D. Valentine and Johnna Bolyard

BOOK REVIEW

Review of *Systems for Instructional Improvement: Creating Coherence from the Classroom to the District Office* — 464
Evthokia Stephanie Saclarides and Kristin E. Harbour

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

Journal for Research in Mathematics Education
2022, Vol. 53, No. 3, 204–226

When Active Learning Is Inequitable: Women's Participation Predicts Gender Inequities in Mathematical Performance

Daniel Reinholz
San Diego State University

Estrella Johnson
Virginia Tech

Christine Andrews-Larson
Florida State University

Amelia Stone-Johnstone
California State University–Fullerton

Jessica Smith
Florida State University

Brooke Mullins
The University of Virginia's College at Wise

Nicholas Fortune
Western Kentucky University

Karen Keene
Embry-Riddle Aeronautical University

Niral Shah
University of Washington–Seattle

This article investigates the implementation of inquiry-oriented instruction in 20 undergraduate mathematics classrooms. In contrast to conventional wisdom that active learning is good for all students, we found gendered performance differences between women and men in the inquiry classes that were not present in a noninquiry comparison sample. Through a secondary analysis of classroom videos, we linked these performance inequities to differences in women's participation rates across classes. Thus, we provide empirical evidence that simply implementing active learning is insufficient, and that the nature of inquiry-oriented classrooms is highly consequential for improving gender equity in mathematics.

Keywords: Grades or audience; Research participants; Students, instructors, or the curriculum in undergraduate mathematics classes; Mathematics, content and processes; Abstract algebra and discrete mathematics; Calculus, analysis, and differential equations; Research areas; Social characteristics; Gender/sexual orientation; Individual characteristics; Identity, agency, authority, and empowerment; Research theories, methodologies, and designs; Mixed methods; Other Topics; Research

Calls for active learning in undergraduate mathematics are now widespread. For instance, the Common Vision Committee—a collaboration among the five largest professional mathematical societies in the United States—summarized this as a need "to move away from the use of traditional lecture as the sole instructional delivery method in undergraduate mathematics courses" and declared that, as a field, "we should seek to more actively engage students than we have in the past" (Saxe & Braddy, 2015, p. 19). This push has been driven by recent empirical evidence demonstrating the positive impact of active learning on student performance across undergraduate science, technology, engineering, and mathematics (STEM) in general (Freeman et al., 2014; Theobald et al., 2020), and mathematics in particular (e.g., Laursen et al., 2014).

Laursen et al. (2014) focused specifically on inquiry-based learning (a particular type of active learning), analyzing data from more than 100 classes from four mathematics departments in which inquiry-based learning had been widely adopted. They found that student performance improved in the inquiry-based learning classes, and when considering students' self-reported experiences, "the use of [inquiry-based learning] eliminates a sizable gender gap that disfavors women

This material is based on work supported by the National Science Foundation under IUSE Awards DUE1431595, DUE1431641, DUE 1431393, and DUE1943146.

students in lecture-based courses" (p. 406). Given the gatekeeping role of undergraduate mathematics courses (Bressoud, 2021), especially for women[1] who disproportionately leave STEM after calculus (Ellis et al., 2016), this was an important and promising finding that supported the recommendation of active learning as a mechanism for both improving student performance and supporting women in undergraduate mathematics. But to what extent are these findings generalizable across institutions and forms of active learning?

Recent studies problematize who, exactly, benefits most from active learning environments. Although the aggregate effects of active learning may be positive, the disaggregated impact on various student groups can still be inequitable—for instance, with boys benefitting more than girls (Bando et al., 2019).[2] Similarly, another recent study on flipped mathematics classrooms found significantly greater performance gains for men and White students compared with women and Black and Latinx students, respectively (Setren et al., 2019). Moreover, because active learning environments increase the frequency of student–student interactions, they can make room for interpersonal microaggressions (Cooper & Brownell, 2016; J. B. Ernest et al., 2019). How do we reconcile the promising findings about inquiry-based learning with these more recent studies showing the potential pitfalls of active learning environments?

To make progress toward reconciling these studies, we focus on two types of equity that are consequential for undergraduate mathematics. First, we consider *participatory equity*, which concerns the fair distribution of opportunities to participate in classroom discourse (Shah & Lewis, 2019). Participatory equity provides insight into whether some groups of students get more opportunities to benefit from particular forms of teaching, such as active learning. Second, we consider *performance equity*, which concerns student performance outcomes. Performance equity would correspond to a condition in which student performance could not be predicted on the basis of student demographics (i.e., performance outcomes are distributed equally across social groups; Gutiérrez, 2002). Student performance is consequential because it dictates whether students pass their courses and have an opportunity to continue toward their STEM aspirations. Ultimately, our analyses aim to connect these two important types of equity, elucidating how particular learning environments may lead to more equitable student performance. Toward this end, we carried out a secondary analysis of 20 undergraduate mathematics classes that were taught using a particular form of active learning called inquiry-oriented instruction. We address the research question: How do talk-based participation patterns in inquiry-oriented classes relate to gendered performance differences in student outcomes?

In this article, we take a critical stance toward active learning, rejecting universalizing ideas that pose active learning as a panacea. Nonetheless, to be clear, we do not advocate against the use of active learning strategies in favor of traditional lecture. Rather, we caution against blanket statements like "active learning is good," and argue that more research is required to better understand how to support instructors' implementation of active learning to ensure that it does not further exacerbate inequities and disproportionately benefit dominant groups in mathematics (e.g., White and Asian men). In this way, active learning may eventually be seen as necessary but insufficient for improving equity in mathematics education.

Prior Research on Inquiry and Gender in Mathematics

"Active learning" is an umbrella term that describes instructional practices focused on actively engaging students in the learning process (e.g., discussion and group work), in contrast to the use of pure lecture (Mathematical Association of America, 2018). Active learning is predicated on the connection between mathematical discourse and learning (Hufferd-Ackles et al., 2004; National Council of Teachers of Mathematics [NCTM], 2014). Beyond supporting learning, public participation offers opportunities for students to be seen as competent mathematicians by others, which supports their mathematics identity development (Langer-Osuna & Esmonde, 2017). In this way, classroom participation is one important mechanism to support students to succeed in mathematics. In active learning classes, a key role for instructors is to create a learning environment in which students have meaningful opportunities to publicly participate in mathematical sense making. In this section, we describe the specific characteristics of inquiry-oriented instruction that may be associated with students' meaningful participation. We also discuss how the gendered nature of mathematics learning environments could be visible in the form of inhibited participation and identity development for women.

Inquiry-Oriented Instruction

The form of inquiry-oriented instruction examined in this study draws heavily from the instructional design heuristics of realistic mathematics education (Freudenthal, 1991). Realistic mathematics education leverages students' informal reasoning as a starting point from which to build more sophisticated mathematical understandings. Inquiry-oriented task sequences,

[1] We recognize that research on gender too often centers White women, or does not attend to race at all, thereby erasing the experiences of women of color. Intersectional frames in mathematics education research are critically important (Bullock, 2018; Leyva, 2017; McGee & Bentley, 2017). As we explain in the methods, our use of "women" in this article includes women across racialized groups as well as transgender women.

[2] This large-scale study drew from 10 field experiments across four Latin American countries (Argentina, Belize, Paraguay, and Peru), consisting of more than 17,000 students in all.

which iterate between phases of inquiry and formalization, are usually carried out in collaborative small groups and whole-class discussions. These task sequences are designed to support students' reinvention of important mathematical ideas. Unlike other instructional approaches, formal mathematics (i.e., definitions and theorems) are not the starting point for the students' mathematical activity (Kuster et al., 2018). Rather, formal mathematics emerges through guided reinvention (Freudenthal, 1991). The bottom-up reinvention of the mathematics, along with the expectation that the students retain ownership over the developing mathematical ideas, requires instructors to skillfully draw out and use student contributions to advance their mathematical agenda (Johnson, 2013; Kuster et al., 2018; C. Rasmussen & Kwon, 2007; Speer & Wagner, 2009).

Whole-class discussions are central to inquiry-oriented instruction. Although private thinking time, peer discussions, and small group work support student thinking, whole-class discussions are the primary mechanism for advancing a collective mathematical agenda. During these discussions, the instructor must elicit student ideas and use them to develop formal mathematical ideas (Kuster et al., 2018; Speer & Wagner, 2009). Which students get to share their thinking, at what depth they share, and what the instructor does with student contributions are all important equity considerations during facilitation. Without careful attention to these issues, gendered classroom dynamics may inhibit women's small-group contributions from becoming public in the whole class discussion (e.g., J. B. Ernest et al., 2019). In addition, these overall dynamics and the social culture of the classroom affect how students value the contributions of their peers (J. Hall et al., 2020). Because whole-class discussions are central to inquiry-oriented instruction, and are susceptible to gender inequities, they were a critical object of focus for the current article.

Gender in Mathematics

Prominent and false cultural narratives circumscribe who can and cannot do mathematics, such as "Asians are good at math" (Shah, 2019; Wu & Battey, 2021) or the "White male math myth" (Stinson, 2008). As one student in a graduate-level mathematics course described, learning proof requires one to "conform to rigid conventions and opinions of elegance as defined by now dead white men" (Reinholz, 2018, p. 71). These studies illuminate how "the popular image of mathematics is that it is difficult, cold, abstract, ultra-rational, important and largely masculine" (P. Ernest, 1993, p. 53). These stereotypes create barriers to success for students with minoritized social identities (e.g., those that are based on disability, gender, race, etc.). Here, we focus on gender.[3]

Gender narratives organize a learning environment by signaling which gender groups do and do not belong in the classroom. For instance, subtle contextual cues (e.g., stereotypical images like Star Trek or video games) can create an environment that reinforces masculine stereotypes, thereby reducing women's sense of ambient belonging and deterring their interest in a masculinized subject domain (Cheryan et al., 2009). Women must also contend with stereotype threat (Steele, 1997). When placed in a situation in which a risk of judgment on the basis of negative stereotypes exists, this risk elicits a disruptive state that inhibits performance (Spencer et al., 1999; Walton & Spencer, 2009). Notably, Black women experience stereotype threat in more harmful ways than White women, because they contend with intersectional forms of oppression (Abdou & Fingerhut, 2014).

Cultural narratives also manifest in classrooms through implicit biases. *Implicit biases* are expectations and evaluations made about groups of people, operating outside of conscious control (Staats et al., 2017). Biases are the result of socialization in particular cultural contexts (Yogeeswaran et al., 2016), and, thus, when people are socialized in environments containing oppressive narratives about women in mathematics, they can develop biases that women are not good at mathematics. These biases—which are ubiquitous in our society—then manifest in concrete ways that marginalize women in mathematics. For instance, teachers underestimate the abilities of young girls in mathematics (Robinson-Cimpian et al., 2014) and call on them less frequently (Sadker et al., 2009).[4]

Classroom spaces are also gendered through student interactions. For example, small groups comprising predominantly men can create a masculinized social environment that inhibits women from participating (Dasgupta et al., 2015). Research also shows that these gendered roles can emerge spontaneously unless instructors explicitly intervene (Quinn et al., 2020). These issues may be further exacerbated when students engage in sexist interpersonal microaggressions (Suárez-Orozco et al., 2015), such as when students in small groups in inquiry-based undergraduate mathematics courses make misogynistic statements implying that "a women's place is in the kitchen" (J. B. Ernest et al., 2019, p. 164). Especially in a highly masculinized environment like a mathematics classroom (Leyva, 2017; Lubienski & Ganley, 2017; Mendick, 2006), these problematic events send strong messages about belonging. These phenomena can be connected to implicit biases about women in general, and particularly about women in mathematics.

[3] We take a fluid view of gender as a field of possible discursive expressions (i.e., many genders exist, not just two; Butler, 1990), as opposed to the historically dominant perspective on gender as a binary.

[4] Similar biases exist toward racially minoritized students (e.g., Larnell et al., 2014; McAfee, 2014; Robinson-Cimpian et al., 2014) and disabled students (Lambert & Tan, 2017).

Of course, the ways in which minoritized gender groups experience oppression is intertwined with oppression related to other social markers (Crenshaw, 1990). Intersectional approaches in mathematics education research have, for example, revealed distinct differences in the learning experiences of girls and women of color (Gholson, 2016). Intersectional oppression toward racially minoritized women is well documented in mathematics education (Joseph, 2017; Joseph et al., 2017; Leyva et al., 2021; Shah et al., 2020). This is important to acknowledge because, too often in research, the category of "women" has wrongly been taken to code for only White women, thereby erasing women of color. In response to this concern, our study sample includes women from a variety of racialized groups. Still, limitations of the secondary data set we analyzed made it difficult to analyze specific race–gender intersections in this particular article; we elaborate on this issue in the Methods section.

The salience of gender in the mathematics classroom has important implications for instruction. Student participation is central to most active learning approaches (including inquiry-oriented instruction). However, because of the high frequency of interpersonal interactions and the possibility those create for marginalizing events (Cooper & Brownell, 2016), not all students will have equitable access to those forms of participation most supportive of learning. In particular, given the importance of student contributions to the whole class in inquiry-oriented instruction, instructors can play an important role in choosing who has an opportunity to present and whether they will be positioned as competent. Unless instructors approach this role intentionally, they are likely to reinforce their own implicit biases and the biases of their students. Given that whole-class discussions are a highly visible form of participation, women being seen as competent in this venue is especially important (J. B. Ernest et al., 2019; Solomon et al., 2011). In addition, if men increasingly see women as publicly successful in mathematics, it has the potential to shift attitudes and stereotypes more broadly.

Author Positionality

The author team consists of mathematics educators who are committed to inquiry-oriented instruction in mathematics. Multiple authors were members of the original Teaching Inquiry-oriented Mathematics: Establishing Supports (TIMES) project, which was important for our understanding of the rationale behind design and data collection decisions made in the original project. Our team also represents a range of intersectional social marker identities by gender (men, women, nonbinary) and by racialized group (Asian, Black, Latinx, White). Broadly speaking, our team is attuned to issues of gender inequity, and also its intersections with other identities such as race or disability. Three women on the team (two White and one Black) had firsthand experiences of gender and intersectional oppression in mathematics classrooms, which was helpful in interpreting gendered segments of classroom interaction in the recorded videos. In discussions about the data and framing of the article, we brought to bear our lived experiences from multiple identities and perspectives to represent the data as comprehensively as possible.

Methods

TIMES was a professional development project that supported instructors who were implementing inquiry-oriented linear algebra (Wawro et al., 2013), inquiry-oriented differential equations (C. Rasmussen et al., 2018), or inquiry-oriented abstract algebra (Larsen et al., 2013). Instructors in TIMES received three forms of support: curricular materials, summer workshops, and online working groups. The inquiry-oriented curricular materials consist of task sequences with rationale, examples of student work, and implementation suggestions. The 3-day summer workshops were designed to help instructors understand the principles of inquiry-oriented instruction (as described earlier) and how to use the instructional support materials. The online working groups, held throughout the semester of implementation, were hour-long weekly meetings with two components: an open forum devoted to addressing issues and concerns for participants as they arose (e.g., challenges of facilitating group work and discussion) and lesson studies (Fortune & Keene, 2021). During the two lesson studies each term, the working group first discussed the mathematics of a focal unit and then discussed learning goals and implementation. After instructors taught the unit, they shared video-recorded clips of their instruction for group reflection and discussion. Throughout the sessions, groups attended to the four components of inquiry-oriented instruction—generating student ways of reasoning, building on student contributions, developing shared understandings, and connecting to standard mathematical language and notation (Kuster et al., 2018).

TIMES Participants

During the 3-year project, a total of 42 instructors participated as TIMES research subjects. For each of these 42 instructors, students were asked to complete content assessments at the end of the term, and classroom video data were recorded during two instructional units. However, the TIMES project team was able to gather gender data from students in only 24 of the 42 instructors' classes, all of which were either differential equations or abstract algebra classes. When investigating performance equity, we draw on the content assessments collected in these 24 classes, as well as comparison data sets

Table 1

Student Demographics for the Inquiry-Oriented Instruction Classroom Sample

Demographic category	Abstract algebra (n = 59[a])	Differential equations (n = 222[b])	Overall (N = 281)
Women	44.1%	30.2%	33.1%
Men	52.5%	67.6%	64.4%
Nonbinary	0.0%	1.4%	1.1%
No answer	3.4%	1.0%	1.4%
White	74.6%	47.7%	53.4%
Latinx	2.0%	14.4%	11.4%
Asian	8.5%	16.2%	14.6%
Black	0.0%	8.1%	6.4%
Native[c]	0.0%	1.0%	1.0%
Multiracial	6.8%	10.4%	9.6%
Unknown	8.5%	2.3%	3.6%

[a] In the nine abstract algebra classes, 122 students consented to data collection. However, only 59 (from three classes) were asked for full demographic data as part of the content assessment.
[b] In the 11 differential equations classes, 288 students consented to data collection, but only 222 provided full demographic data as part of the content assessment.
[c] The Native category was a combination of two demographic categories, American Indian or Alaskan Native and Native Hawaiian or other Pacific Islander.

collected from instructors that were not part of the TIMES professional development program (see more about the content assessments and the comparison data in the Data Sources section).

When collected, questions about student demographic data were included as part of the content assessments. This allowed us to disaggregate student performance data by student demographics. However, because these content assessment data were anonymized, we could not connect student performance and classroom participation at an individual level, and instead had to focus on gendered analyses at a classroom level (a point we return to later when we discuss the video data). Because we wanted to investigate the possible relationships between performance and participatory equity, we removed four instructors from the sample for whom student assessment data was limited (i.e., only a single performance assessment for either women or men). Thus, for data and analysis presented at the classroom level, this left us with N = 20 instructors total (13 women, seven men; race data were not available). These 20 participants taught at different universities across the United States. The sample consisted of a mix of predominantly White institutions and Hispanic-serving institutions.[5]

Although our analyses focus on gender only, we provide the available racial demographics in Table 1 to help contextualize our findings and their potential generalizability (with more details presented in Appendix A).[6] Whereas student gender and race data were collected in all 11 inquiry-oriented differential equations classes, no race data were collected for students in inquiry-oriented abstract algebra during the first year of data collection. Thus, whereas we have gender data for all nine abstract algebra classes, we have race data in only three. For this more limited set of students, we note that the distribution of gender and race was not uniform across abstract algebra and differential equations.

Data Sources

A variety of data were collected from the instructors using inquiry-oriented instruction. Relevant to this article are student content assessments and video recordings of instructional units.

Content Assessments

We used the Group Theory Content Assessment to measure conceptual understanding of abstract algebra (Melhuish, 2015). This instrument was developed using a mix of classical test theory and item-response theory (Melhuish, 2019). For the abstract algebra comparison (noninquiry) group in the current article, we analyzed data from the final round of

[5] We use the term “Hispanic-serving institutions” as a federal designation (Garcia, 2017), yet simultaneously recognize the contested nature of language related to groups broadly categorized as “Hispanic,” who may prefer Latino/a/x, Chicano/a/x, or other terms depending on their own racial/ethnic identities and identification (de Onís, 2017).

[6] Disaggregated demographic data are provided in Appendix A. Table 1 gives an overview summary of race and gender separately.

the instrument development process. This group contained 374 students across 33 institutions with varying levels of selectivity ranging from most selective (<25% of students accepted) to least selective (100% of students accepted). The 12 inquiry-oriented abstract algebra instructors[7] taught a total of 174 students. Of those students, 139 (80%) completed the Group Theory Content Assessment. In total (i.e., students in inquiry-oriented and noninquiry classes), our sample contained 269 men, 237 women, and seven who identified as nonbinary or declined to answer. The gender makeup was very similar between the noninquiry and inquiry-oriented data sets (47.6% and 42.5%, respectively, identified as women).

For the differential equations participants, TIMES project personnel developed a differential equations content assessment, assessing both conceptual and procedural knowledge (W. Hall et al., 2016). Each TIMES instructor collected data from all consenting students in their classes. Additionally, every instructor identified a colleague at their same institution who was also teaching differential equations but not using the TIMES materials or inquiry-oriented instruction.[8] Because our participatory analyses are unable to account for race, we report only gender data in the noninquiry sample. In all, 448 students took the differential equations content assessment. However, only 341 of those 448 were asked for their gender. In total (i.e., students in inquiry-oriented and noninquiry classes), our sample contained 225 men, 101 women, and 15 students who identified as nonbinary or declined to answer. The gender makeup was nearly identical between the noninquiry and inquiry-oriented data sets (29.7% and 29.6%, respectively, identified as women).

Classroom Video Data

We had access to two recorded units of instruction for each instructor. Our analyses focused on the second unit (M = 145.4 min in duration), recorded 8–10 weeks into the semester, because it was most likely to represent established participation patterns in a given classroom. This unit was typically taught during two to three class periods. For all the abstract algebra instructors, this was a unit on quotient groups. For the differential equations instructors, each online working group determined which units they wanted to focus on during the lesson studies, so not all the differential equations videos were of the same unit. Videos were recorded on a tablet in the back of the room, which captured whole-class discussions but did not allow us to follow small group interactions. Thus, our analyses focus on coding only whole-class discussions.

The original data corpus did not include a class roster or seating map of the students, and because assessment data were anonymized, we could not connect reported demographics from the assessments to the students in the videos. Therefore, for this article, we operationalized gender in the classroom video data through gender performance (M. L. Rasmussen, 2009). That is, students' gender was inferred using visual and audio cues (e.g., voice, clothing, presentation, names, or pronouns used) by three members of the team (all women: two White, one Black). Although three students in the inquiry-oriented sample identified as nonbinary and four students declined to answer the question about gender, we did not encounter gender-neutral pronouns in any of the videos. We had no other way to infer whether students were nonbinary, so we cannot determine whether participation from any of the seven nonbinary students was captured on video and the students were misgendered by our coders, or whether their participation was simply not captured. Hence, our claims are limited to binary interpretations of gender performance. Overall, when we could not identify the gender of a particular speaker, their contribution was not coded.

Despite this limitation, a theoretical assumption of our work is that gendered social interactions are largely guided by the gender expression perceived by students' instructors and classmates. This assumption is consistent with studies of implicit bias, which typically focus on perceived gender, rather than actual gender identity (Yogeeswaran et al., 2016). Such work suggests that an individual's perceptions of others will be the predominant factor affecting how biases play out. Similarly, work in queer studies suggests that gender expression (more than identity) is a significant factor in predicting bias and discrimination toward women (Levitt et al., 2012). Simply put, when biases play out in social contexts, the effects of those biases tend to mirror the social marker identities that we implicitly ascribe to others. Therefore, the cues available to teachers and students in the classroom that would influence perceived gender are the same cues that we researchers used when analyzing the classroom data.

Although we were interested in exploring intersections between gender and other social markers, we did not pursue this angle because of a lack of relevant context clues. Whereas students were referred to by their gender pronouns, we had no equivalent marker for racial identities, for example. Drawing inferences about race based solely on phenotype would have been problematic, and given our secondary analyses of video data, other potentially relevant contextual data were not always present (e.g., a student's name).

[7] Note that three of these classes were later dropped from the analysis, but this did not have a significant impact on the proportion of women in the sample.

[8] One exception was a participant from a small university who instead recommended a colleague from a similarly sized university located in the same city.

Affordances and Limitations of Secondary Data Set

Performing a secondary analysis of existing data led to additional complexities in the study. In particular, we had to coordinate multiple data streams (i.e., student performance assessments and classroom observations) that were not linked in the initial study, which prevented us from performing some types of analysis (e.g., intersectional analyses of race and gender). Had we designed the study from the outset, the analytic path would have been much more straightforward. However, these limitations were outweighed by the affordances of the sample. In particular, we had a large data corpus describing inquiry-oriented classrooms taught by instructors who had received considerable support. Moreover, these inquiry-oriented classrooms had gendered inequities in student outcomes that did not exist in the noninquiry comparison sample. Thus, the goal of our secondary analyses was to shed light on these important—and counterintuitive—findings. This work also positions future researchers to take up questions of classroom participation and gendered performance more directly through intentional study design.

Analytical Methods

Broadly speaking, we were guided by the equity analytics methodology (Reinholz & Shah, 2018). This approach focuses on identifying and analyzing quantitative patterns in student participation as a window into classroom equity. A central idea behind equity analytics is that quantitative data themselves cannot say whether participatory equity has been achieved in a classroom because determinations of equity and inequity also require a deep understanding of students' subjective experiences. Instead, analyses of quantitative data can provide insight into statistical in/equalities, which serve as a waypoint toward determinations of in/equity. In particular, we argue that students who are minoritized in a discipline (e.g., women in mathematics) should receive at least a proportional share of participation opportunities. Anything less would be considered a sign of inequity. Drawing from this broad conceptualization of equitable participation, we aimed to link inequities in participation patterns to inequities in student performance.

Variable Construction

In this study, we needed a way to analyze both student performance and student participation. To do so, we constructed performance and participation variables that would account for student gender while not masking differences between the two content areas and differences among classes (e.g., percentage of women students).

Performance Measures. The content assessments were collected with two different instruments in two different types of courses. Therefore, we transformed the data to allow for comparison. First, the distributions and averages of the scores between the two content areas were not comparable, meaning that (for instance) a 65% on the differential equations assessment had a different meaning than a 65% on the abstract algebra assessment. To enable us to compare scores across the two assessments, we converted these percentages to *z*-scores.

Second, the two content areas displayed differences in gendered performance on the content assessment. In the noninquiry abstract algebra data set, women averaged 3.3% lower than men (or −0.17 *SD*), whereas in noninquiry differential equations data set, women scored 3.5% higher than men (or 0.21 *SD*). These average differences became the respective zeros for the inquiry-oriented abstract algebra and inquiry-oriented differential equations gendered performance differences. Thus, the gendered performance difference in each individual inquiry-oriented class was computed by subtracting these national comparison differences from its *z*-scores. These two steps result in the creation of a variable for each inquiry-oriented teacher that took into account difference in the content assessments and in the course subjects.

Student Participation Measure. We used the classroom observation tool EQUIP (Equity QUantified In Participation) to quantify patterns of student participation (Reinholz & Shah, 2018; https://www.equip.ninja). These data draw attention to how different groups of students in the same classroom (e.g., by gender) may have different access to meaningful participation and learning through classroom discourse. In this article we used a total of five dimensions customized to capture inquiry-oriented discourse (see Appendix C). We used generic EQUIP codes for teacher question, student talk type, and student talk length. We customized solicitation method and teacher response to better capture features of inquiry-oriented instruction. In particular, we added additional levels of the codes to distinguish when individuals, groups, or volunteers were called on. We also customized teacher response to capture teacher Elaboration, Revoicing, and Follow Ups. This allowed us to explore greater detail than the original EQUIP codes, which focused only on whether a teacher evaluated student ideas.

In EQUIP, the unit of analysis is a *student contribution* (originally "participation sequence"—see Reinholz & Shah, 2018), which starts when a new student speaks and ends when another student speaks. With this definition, any length of interaction between the instructor and student is coded as one contribution. Conversely, if two students are having a conversation, then a new contribution begins each time a student speaks, resulting in many back-to-back lines of code.

Table 2

Intercoder Reliability (Krippendorff's Alpha) Between Double Coders and the Lead Coder

Coder	Number of double-coded sequences	Solicitation method	Length of talk	Student talk type	Instructor solicitation	Instructor response
1	336	.97	.97	.85	.89	.80
2	470	.98	.98	.85	.89	.82

Note. Krippendorff's alpha > 0.8 indicates a high level of reliability.

Thus, by the design of EQUIP, in classrooms with more interactions among different students (rather than a single student and the teacher), the number of contributions coded will be higher. This design feature of EQUIP becomes an important factor in interpreting our quantitative results, and we will return to it in the Results section.

The coding team consisted of three graduate students (all women; two White, one Black), with one lead coder who had extensive prior experience with EQUIP. To begin, a high degree of intercoder reliability was achieved over approximately 20% of the data set (21 hr of video total); Krippendorff's alpha is given in Table 2. All results were more than .8, the highest category that can be achieved (Carletta, 1996). Once intercoder reliability was established, each graduate researcher coded approximately one-third of the data set that was randomly assigned to them.

Quantitative Analysis

Our quantitative analysis had five main phases: (a) content assessment analysis, (b) identification of descriptive features of inquiry-oriented class discussions, (c) regression analysis relating participation to performance among women, (d) identification of groups of classes with gendered performance differences, and (e) analysis of differences among groups of classes with gendered performance differences. In our first phase, we began by preparing the data set for analysis. First, we examined gendered performance differences between inquiry-oriented and noninquiry classes to see how this instructional approach related to performance differently for men and women.

In our second phase, we examined the features of classroom discussions in inquiry-oriented sections using EQUIP to better understand, using a quantitative lens, typical features of whole-class discussion in inquiry-oriented classrooms. At this point, we dropped the four instructors from the data set who had limited assessment data. This reduced noise in our sample when linking participation and performance (for $N = 20$ instructors total).[9] We also scaled all participation measures by the length of the average recorded unit (145.4 min, or approximately two 75-min sessions), so that we could meaningfully compare across classes for which the length of the coded unit was different.

During the third phase, we performed an extensive exploratory analysis to look for differences between the inquiry-oriented abstract algebra and differential equations classes, given the differences in course content and student demographics. However, we found minimal differences between the two types of courses in observed differences in participation, so we continued instead to work with the data set as a whole. To account for the relative proportions of women and men in each classroom, we created a new measure called a participation rate. The *participation rate* for women (or men) in a classroom is defined as the total number of contributions from women (or men) divided by the number of women (or men) in the classroom. This allowed us to consider the average rates of participation for gender groups. For example, if 50 contributions were made by women in a particular classroom that had five women, we would calculate women's participation rate as 10. When we give means and standard deviations of student participation, we are providing metrics about participation rates.

Next, we used our classroom-level variables (i.e., participation rates and the gendered performance difference) to construct weighted regression models predicting differences in women's and men's performance on the basis of their participation. Because we had some small class sizes in the sample, we weighted the regression equations with $1/SE^2$ (see Appendix B), so that classes with more reliable estimates (i.e., lower standard error) would be weighted more heavily in the models.[10] Through our analyses, we found that correlations between performance and participation rates for men ($r = -.11$, $p = .7$) and women ($r = .54$, $p = .01$) were in different directions, and only women's participation rate was significant, so we conducted follow-up analysis focused solely on women's participation. For simplicity, we then constructed a weighted regression (in R©), with women's participation rates as the only predictor of gendered performance differences. We included instructor gender and a variety of other covariates in our initial models, but none was significant. Although

[9] In addition to running the analyses with 20 instructors, we also ran all the analyses with these four instructors included ($N = 24$), and the general trends and findings were the same.

[10] We also created unweighted regression models, and the same general trends held.

we explored the use of other predictor variables in the models (such as women's Why contributions), many of the predictors were highly correlated, so we could not include them in the same model because of multicollinearity.

In our fourth phase, we sorted the classrooms in the data set into three groups centered on the median value of the performance measure (i.e., the gendered performance difference). Our goal with the tertile split was to identify different variations in the style of inquiry-oriented instruction that would be obscured by looking at the classrooms in the sample only along a single continuum. We recognize the ongoing debate about using analysis of variance (ANOVA)-based approaches versus continuous-variable analyses (Iacobucci et al., 2015), which is why we opted to do both. Our rationale was that in our third phase, as we describe in the results, the regression model demonstrated a significant relationship between women's participation rates and gendered performance differences. However, this finding does not fully illustrate the nature of the various inquiry-oriented classes we observed. These differences became easier to interpret upon splitting the classes into three groups.

During the fifth phase, we statistically explored observed differences in the three groups using one-way ANOVAs. We focused on women's participation rate, women's Why participation rate, and women's Not Called On participation rate, because these were the dimensions that we hypothesized would be most related to student performance according to prior research (e.g., Chi et al., 1994; Engle, 2012). These dimensions provide insight, respectively, into women's overall presence in the discussion; women's opportunities to participate in meaningful, high-level discourse; and women's agency in participating without being directly called on. All these are theorized to be highly consequential for gender equity (Reinholz & Shah, 2018). In particular, prior research shows that although women may make substantive contributions to small-group work, the environment may prevent them from fully participating in whole-class discussions (J. B. Ernest et al., 2019), so insight into women's agency and opportunities to participate was important to capture.

Qualitative Vignettes

After quantifying differences among the three groups of inquiry-oriented classes, we selected three vignettes to illustrate the nature of participation in each group. Each vignette was chosen to be illustrative, but not necessarily representative, of the quantitative, statistical differences observed among the groups. As we introduce the vignettes in our results, we describe in depth how the particular features in the vignette illustrate the quantitative EQUIP coding. These vignettes are not intended to provide a full qualitative analysis but rather illustrate some potential explanations of the quantitative results. The vignettes help the reader get a feel for what the inquiry-oriented classrooms looked like, and how different instructional styles could lead to gendered performance differences.

Quantitative Results

Student Performance Outcomes

We found no significant differences in performance for women between inquiry-oriented and noninquiry classes (see Table 3). In contrast, men in the inquiry-oriented classes significantly outperformed men in the noninquiry classes. The results were significant for both differential equations (55.1 vs. 47.7, $p = .002$) and abstract algebra (50.4 vs. 42.9, $p = .007$). This improvement for men in abstract algebra coincides with a statistically significant gendered performance difference between men and women in inquiry-oriented abstract algebra ($p < .001$). A full analysis of this finding, using hierarchical linear modeling, can be found in Johnson et al. (2020). Thus, even though women in the inquiry-oriented classes did no

Table 3

Average Student Performance on Conceptual Assessments for Inquiry-Oriented and Noninquiry Classes (Raw Scores out of 100)

Category	Inquiry-oriented instruction	Noninquiry	Independent samples t-test
	Differential equations		
All students	55.0 ($n = 230$)	49.0 ($n = 111$)	$t(339) = 3.2^{**}$
Women	54.6 ($n = 68$)	51.2 ($n = 33$)	$t(99) = 1.04$
Men	55.1 ($n = 151$)	47.7 ($n = 74$)	$t(223) = 3.2^{**}$
	Abstract algebra		
All students	44.4 ($n = 139$)	41.4 ($n = 374$)	$t(511) = 1.5$
Women	35.7 ($n = 59$)	39.6 ($n = 178$)	$t(235) = -1.4$
Men	50.4 ($n = 77$)	42.9 ($n = 192$)	$t(267) = 2.7^{**}$

$^{**}p < .01$.

worse than women in noninquiry classes, the significant improvement for men in inquiry-oriented classes resulted in an overall gender inequity. Across the sample of 20 classrooms analyzed, only five had a gendered performance difference with women outperforming men; men outperformed women in the other 15 classes (see Appendix B).

Student Participation

Summary of Participation

Table 4 summarizes student participation across the entire sample of inquiry-oriented classes. These classes were relatively small (approximately six women and 11 men on average), and discourse-heavy (with participation rates of approximately 4.0 for women, and 6.0 for men, per 145-min unit). Not Called On was the primary form of solicitation (with a participation rate of 2.9 for women and 4.4 for men), which indicates that students often contributed to discussions in a free-flowing way in which they were not explicitly called on by the instructor. The student talk type of What questions, which focus on factual statements, were the primary form of contribution. Moreover, students tended to give elaborate responses (five or more words), and instructors used a variety of discourse moves to build on student thinking (e.g., Revoicing, Elaborating, or asking a Follow-Up question). In general, these findings support the conclusion that instructors were authentically implementing

Table 4

Mean (Standard Deviation) Participation Rates by Gender and Classroom (N = 20)

Dimension	Women	Men
Number of students	6.1 (3.7)	11.1 (6.4)
Contributions	4.0 (3.4)	6.0 (6.0)
	Solicitation method	
Group	0.08 (0.1)	0.19 (0.3)
Individual	0.6 (0.6)	0.7 (0.7)
Volunteer	0.4 (0.5)	0.8 (1.0)
Random	0.01 (0.1)	0.0 (0.01)
Not Called On	2.9 (3.4)	4.4 (4.9)
	Instructor solicitation	
Why	0.4 (0.3)	0.8 (1.1)
How	0.2 (0.2)	0.3 (0.4)
What	1.7 (1.6)	2.5 (1.1)
Other	0.6 (0.5)	0.8 (0.6)
N/A	1.2 (1.6)	1.7 (1.8)
	Student talk type	
Why	0.5 (0.5)	0.7 (0.5)
How	0.2 (0.3)	0.3 (0.3)
What	2.6 (2.5)	4.2 (4.9)
Other	0.9 (0.7)	0.9 (1.1)
	Student talk length	
21+ words	1.0 (0.8)	1.7 (1.3)
5–20 words	2.1 (2.0)	3.1 (3.9)
1–4 words	0.9 (1.0)	1.2 (1.2)
	Instructor response	
Elaborate	1.2 (1.4)	1.6 (1.6)
Revoice	0.6 (0.8)	1.0 (1.3)
Evaluate	0.4 (0.8)	0.3 (0.4)
Follow-Up	0.6 (0.8)	0.9 (1.7)
N/A	1.2 (1.0)	2.2 (2.1)

inquiry-oriented instruction in their classes across the sample (as defined in Kuster et al., 2019). Moreover, these descriptive findings show that, in the aggregate, the nature of participation patterns looked relatively similar for men and women—except that men had more contributions, especially of the Not Called On nature (4.4 for men vs. 2.9 for women).

Predicting Differences in Performance Outcomes by Participation

Given the variation in the amount and types of participation, as indicated in Table 4, we constructed a regression model to see whether differences in women's participation rates (the total contributions from women divided by the total number of women in a class) across classes could be used to predict gendered performance differences (see Table 5 and Figure 1). We did not include men's participation in the model, because participation rates and performance outcomes were only significantly correlated for women, not men.

We found that women's participation rates were a significant predictor of gendered performance differences. Table 5 indicates that the final regression equation was:

$$y = a + Bx$$
$$y = -0.59 + 0.09x,$$

where a is the intercept, B *is slope coefficient*, y is the gendered performance difference, and x is women's participation rate. The intercept of −0.59 in our model indicates that if women never participated, their scores would be roughly half a standard deviation lower than the gendered performance difference in a typical noninquiry class. The coefficient of 0.09 indicates that a participation rate of approximately 11 for women ($0.09 \times 11 \approx 1$) would reduce the gendered performance difference by one standard deviation. Considering the average of approximately six women per classroom, this would require increasing women's participation by 66 contributions total over the course of the 145-min unit, which is a feasible goal. This model is illustrated graphically in Figure 1.

Table 5

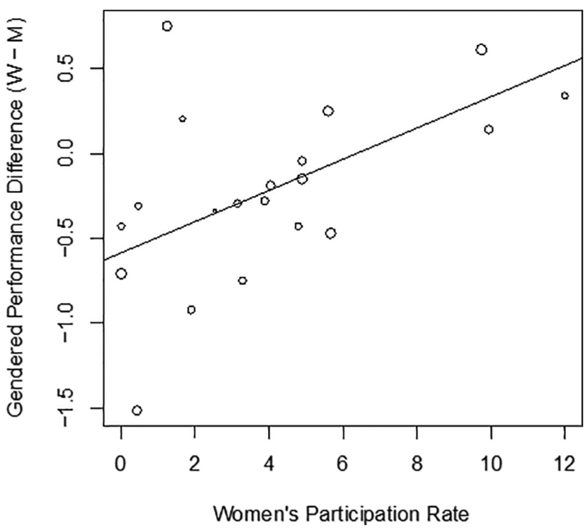
Weighted Multiple Regression (Gendered Performance Difference Between Women and Men, N = 20 Classrooms)

	B (unstand.)	*SE B*	β (Stand.)	*t*
(Intercept)	−0.59	0.18	0	−3.2**
Women's participation rate	0.09	0.03	.59	2.7*

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

Figure 1

Weighted Regression (Gendered Performance Difference Predicted by Women's Participation Rate)



Note. The size of each circle is scaled by $1/SE^2$, showing the weighting factor in the regression equation.

Table 6

Mean (Standard Deviation) Participation Rates by Group

Category	Group 1: centralized ($n = 6$)	Group 2: distributed ($n = 8$)	Group 3: shared ($n = 6$)	Two-way ANOVA
Participation rate (all students)	2.0 (1.5)	4.7 (1.0)	9.5 (6.8)	$F(2, 17) = 5.9^{*}$ Tukey HSD[a]: Group3−Group1**
Women's contributions	1.0 (1.3)	4.2 (1.0)	6.7 (4.6)	$F(2, 17) = 6.9^{**}$ Tukey HSD: Group3−Group1**
Women's Why contributions	0.2 (0.3)	0.5 (0.3)	0.7 (0.7)	$F(2, 17) = 2.5$
Women Not Called On	0.6 (0.9)	2.4 (0.7)	5.8 (5.0)	$F(2, 17) = 5.5^{*}$ Tukey HSD: Group3−Group1*
GPD[b] (W−M)	−0.9 (0.3)	−0.3 (0.1)	0.3 (0.3)	
Number of students	22.6 (7.0)	14.1 (7.0)	15.5 (7.7)	
%Women	27.0	43.0	39.0	

[a]Tukey's honest significant difference.
[b]GPD = gendered performance difference.
$^{*}p < .05$. $^{**}p < .01$.

Further Exploring Differences in Participation by Inquiry Type

Once we established the significant relationship between women's participation and the gendered performance difference, we sorted the classrooms in the data set into three groups to further explore the relationship between participation and gendered performance differences (c.f. Appendix B). A quantitative summary for the classrooms in each of the three groups is given in Table 6, which shows significant differences in how women participated in each of these groups.[11] These findings led us to name the groups Centralized, Distributed, and Shared Inquiry, as we elaborate next.

We found significant differences in participation rates across groups. The classes in which women had the lowest performance outcomes compared with men (Group 1) had significantly lower participation rates for women. They also tended to have the fewest contributions overall and fewer women in the classes. We called Group 1 the *Centralized Inquiry* group because the relatively lower number of contributions can be explained by the way that contributions are defined in EQUIP. Namely, this indicates that a high proportion of the discourse flowed back and forth between the instructor and a single student at a time (i.e., talk was centralized through the instructor).[12] This does not mean that Group 1 used lecture—on the contrary, these classrooms had many in-depth discussions—but that the instructor was at the center of these discussions.

We named Group 2 *Distributed Inquiry* because the participation rates for women were closest to the overall participation rates for all students. A rate of 4.2 for women was just slightly below the overall rate of 4.7 (and 5.2 for men), whereas in the Group 3 classes, the rate of 6.7 for women was considerably lower than the overall rate of 9.5 (and 10.5 for men). Thus, in these Group 2 classes, the ratio of participation was closest to 50–50 between men and women. Note that whereas in Group 2, women participated more relative to men, the greater overall amount of discourse in Group 3 resulted in higher participation rates for women.

In Group 3 classes, women had the highest assessment scores compared with men. We found women in Group 3 classes had significantly higher participation rates, in general and when Not Called On, compared with Group 1. Although we also found numerical differences in Why contributions between groups, we could not detect a statistical difference. We called Group 3 the *Shared Inquiry* group because the greater quantities of talk—and particularly of Not Called On talk—indicate that students were responding to one another, sharing in the inquiry process (instead of interacting primarily through the instructor).

[11] Here, and other places in this report, we focus on women's participation because that has been found to be highly correlated with their performance, whereas men's participation was not, as described in the Methods.

[12] As we discussed in the Methods section, when talk flows back and forth between a single student and the instructor, it is coded as a single contribution. If the same conversation took place between two students, the overall number of contributions would increase. Because instructors were using the same curriculum, the relative amount of whole-class discourse was relatively uniform among classes. Thus, the primary driver of differences in overall participation rates was the nature of the discourse.

Summary of Quantitative Results

Our quantitative analyses established a few important empirical results. First, we found that significant gendered performance differences existed between women and men in the inquiry-oriented sample that were not present in the noninquiry comparison sample. Second, we linked these performance differences to women's participation rates in the different classes. In other words, we could predict in which classes women would perform best by looking at in which classes they participated the most. Third, we found some quantitative evidence that the nature of participation, not simply the amount, was consequential to gender equity. In the qualitative vignettes in the next section, we provide a further illustration of how the nature of inquiry-oriented instruction differed between classes.

Qualitative Vignettes

The quantitative findings provided evidence that instructors were authentically implementing inquiry-oriented instruction, not teaching through pure lecture. However, differences in the nature of participation across classes suggest that the forms of learning opportunities available differed by class. Here, we offer three brief vignettes to qualitatively illustrate the quantitative differences we found across the groupings of inquiry classes, as described earlier. We chose these vignettes by searching through the data corpus and picking excerpts that exemplified the relationship between the EQUIP coding scheme and the qualitative data (e.g., in Centralized Inquiry we found an example of meaningful back and forth between the instructor and a single student, which was emblematic of these classrooms that had a low number of overall contributions). These vignettes are intended to provide a general sense of the classrooms for readers who are less familiar with EQUIP. We do not claim that these interactions represent all discourse within a given classroom. Rather, we argue that they are emblematic of the types of interactions that happened within their respective group of classes, and thus we offer them for illustrative purposes. All names are pseudonyms.

Centralized Inquiry (Example From Classroom AA1)

The Centralized Inquiry group exhibited significantly fewer student contributions, which was driven in part by a higher frequency of back and forth between the instructor and a single student, rather than student-to-student interactions. A typical example of this discourse pattern consists of the instructor going back and forth with a single student to press them for their thinking. Consider the following episode between the instructor in Classroom AA1 and the student Jamie, in which the instructor pushes Jamie to clarify her thinking:

1. *Jamie* Could we start with the statement, *a* star *x* equals *b* star *y*?
2. *Instructor* OK, so Jamie got another idea. She started with this equation . . . and what did you do with it?
3. *Jamie* Well, we know that all of these elements are elements of group *D*. So, we know that we can take the right inverse of *x* on both sides.
4. *Instructor* So you multiplied both sides by the inverse.
5. *Jamie* And also, groups are associative so we can, yeah.
6. *Instructor* So, what was your next step?
7. *Jamie* Well, we know that we can move the parenthesis, because groups are associative, or elements of groups are associative.
8. *Instructor* Elements are never associative. It's the operation.

In this episode, we see the instructor in a back-and-forth dialogue with a single student in the class to elicit deeper thinking. In this episode, we do see a genuine effort by the instructor to probe into Jamie's thinking. We see teacher discursive moves featuring How questions (Lines 2 and 6), a Revoicing of a student contribution (Line 4), and an Evaluative comment about the student's contribution (Line 8). This exchange is illustrative of this instructor's central role in eliciting and validating the mathematical ideas proposed. Exchanges of this nature were common in AA1, which had only 38 coded contributions over the scaled 145-min unit because much of the discourse occurred between the instructor and a single student, which was coded as a single contribution.

Distributed Inquiry (Example From Classroom DE5)

The Distributed Inquiry group had close to 50–50 participation between women and men and, numerically, the overall participation rates and Not Called On participation rates were in the middle of the other two groups of classes. In Distributed Inquiry classes, the instructor still played a role in mediating the discourse through a mix of Called On and Not Called On interactions. However, these interactions focused more on Distributing participation opportunities across students, rather than back and forth with a single student. The following excerpt is an example of this discourse pattern from Classroom DE5 in which explicit instructor moves (like calling on particular students in an intentional order) elicited a variety of student viewpoints:

1.	*Lisa*	Up until that point, the spring force was accelerating the mass to that point. And then right after that point, the spring force is slowing down the mass.
2.	*Instructor*	So there's no spring force here. . . . Kathy, what do you think about that? I know you have some strong views about that.
3.	*Kathy*	Um so . . . the spring is at a resting point at *x* equals 0. So, if you stretch it out, it will want to compress back to that point. So it's going to accelerate there . . .
4.	*Instructor*	Say that again.
5.	*Kathy*	Like, if it's accelerating to *x* equals 0, so the velocity is increasing towards *x* equals 0, then velocity has to be a maximum at *x* equals 0 for going that same motion.
. . .		
6.	*Instructor*	How about this group up front? You had a different picture, so can you tell me about that?
7.	*Gilbert*	I just drew a few circles because the question asked for a spring . . .

In two instances in this episode, the instructor clearly uses what she saw during prior work time to bring students into the discussion. For instance, the instructor alludes to Kathy's "strong views" on the problem (in Line 2) and also brings in the "group up front," because they had a "different picture" (in Line 6). Here the instructor is using her knowledge of what happened in small groups to orchestrate the whole-class discussion.

The instructor also says, "Say that again" (in Line 4), presumably to ensure that the rest of the class can hear Kathy's idea and respond to it. At this point of the discussion, the instructor only elicits a variety of ideas and withholds her judgment about what is correct. As we can see, the primary role of the instructor in this excerpt was to elicit student thinking, and she ensured that productive conversations were taking place by carefully choosing which students would share their ideas. In our coding, this dynamic is reflected in the relatively higher participation rates, but still modest number of Not Called On contributions. Using these discourse moves, the instructor supported inquiry that was Distributed among the students as they made sense of one another's ideas. These intentional instructor moves also set up students to respond directly to one another. This was evident in the larger number of total contributions in Classroom DE5 than what we saw in AA1: 107 over the scaled 145-min unit.

Shared Inquiry (Example From Classroom DE11)

In the Shared Inquiry group, women's participation rates were the greatest, as were the rates for Not Called On contributions, which occurred when students were collaboratively building on one another's ideas. One of the primary roles for the instructor in this type of discussion was to establish strong community standards for participation and to support students to discuss with one another. An example of this type of discourse can be seen in the following debrief conversation from Classroom DE11. The students were trying to understand the path of a point on a propeller on an airplane that was flying:

1.	*Instructor*	We need to conclude this in the next five minutes. Candy, you start. What were you showing? Sorry I interrupted you.
2.	*Candy*	We were arguing since you're moving with it . . . you should just see a straight line. And I was like, no, because it's still like the plane is still moving. Like whether you like it or not, it's still gonna be moving in this way and the propeller's still gonna be moving in this way. So it's gonna go like this. . . . From the side it's 2D, it's gonna look like a sine wave.
3.	*Instructor*	It's like a sine wave.
4.	*Alyssa*	But, what if we're looking at it from the top though, because . . .
5.	*Candy*	Same thing. Literally just a spiral. So, like, it's like you see a spiral and does it look different if you look at it this way or this way? No, it's just a spiral.
[James raises his hand]		
6.	*Instructor*	OK, James, what do you think?
7.	*James*	I don't think we (inaudible). It says like at a slow constant speed . . .
8.	*Olivia*	Well it says you can always see the red mark. So you're gonna see it.

9. *Candy* It would just be like this. The plane is still moving.
10. *James* Like if you're just standing in place then it looks like a sine wave, but if you're moving along then it's a line.
11. *Instructor* So what's the conclusion, from the top, what do you see?

In this episode, the instructor initially sets up the students to discuss the problem by saying "we need to conclude this." The way that students take this statement as an invitation to arrive at a consensus around the particular problem provides evidence of community standards already established in the classroom environment. Through this episode, this instructor plays the role of a facilitator but does not offer new ideas. The first example is Revoicing Candy's idea that it "looks like a sine wave" (Line 3). Then, the instructor calls on James (Line 6) when he raises his hand. Finally, in Line 11, the instructor asks, "What's the conclusion?" to try to see what consensus the students were building to. All these instructor moves were used to help the students engage in Shared Inquiry. In contrast to Distributed Inquiry, the instructor played less of an intentional role in selecting which students would participate (i.e., more Not Called On talk). Because of this dynamic, Classroom DE11 had an even greater number of total contributions than the previous two: 144 over the scaled 145-min unit.

Discussion

The goal of this article was to understand how talk-based participation patterns in inquiry-oriented classes relate to gendered performance differences in student outcomes. Our focus on participation was guided by theories suggesting that participation is highly consequential both for learning (Hufferd-Ackles et al., 2004; NCTM, 2014) and identity development (Langer-Osuna & Esmonde, 2017). From these perspectives, one would suspect that students' participation would be empirically linked to their performance outcomes, but few studies to date have documented this link (for notable exceptions, see Banes et al., 2020; O'Connor et al., 2017). By providing further evidence of such a link as it relates to gendered performance differences, this article makes a notable contribution to the field.

Contrary to the conventional wisdom that active learning promotes equity, across a sample of 20 undergraduate mathematics classrooms, we found evidence of greater gender inequity in favor of men in the inquiry-oriented instructional environments more often than not (in 15 of 20 classes). This result extended prior analyses that looked at gendered performance differences in a subset of the classes (inquiry-oriented abstract algebra) using hierarchical linear modeling (Johnson et al., 2020). The findings are notable given that no significant gendered performance differences existed in the noninquiry comparison samples.

Quantitative analyses showed that these performance differences were significantly related to women's participation rates in the inquiry classes. In particular, we found that, on a classroom level, women's participation rates were highly correlated with their performance ($r = .54$, $p = .01$), whereas men's participation rates were not correlated with their performance. This suggests that women's participation was highly related to their success, whereas men's participation overall was not related to their success. We interpret this finding with respect to the gendered nature of the mathematics classroom, which can be hostile toward women and nonbinary students (J. B. Ernest et al., 2019; Leyva, 2017; Lubienski & Ganley, 2017). In the classrooms exhibiting high levels of participation from women, we suspect that the instructor was able to create an environment that supported women's participation and, in turn, the relatively high levels of participation from women sent further signals that promoted belonging for other women in those classrooms. All of this may have resulted in improved performance. In contrast, men do not regularly face gender identity threats in a mathematics classroom, so regardless of whether high levels of participation from men were present, they could still assume that they belonged. We recognize that these feelings may not be true for all members of gender groups in a classroom, but suggest that they describe general trends across gender groups.

Given the relationship between participation rates and performance for women, we focused our analysis on women's participation. Through a weighted regression analysis, we found that women's participation rates were a significant predictor of gendered performance differences. In particular, we found that an increase in women's participation rate by approximately 11 contributions during a 145-min unit corresponded with a decrease in the gendered performance difference between women and men by one standard deviation. For a class with six women, this means that an increase of about 66 contributions for women would correspond to a decrease in the gendered performance difference by one standard deviation. Because Shared Inquiry classes had, on average, approximately 85 more contributions than Centralized Inquiry classes, significantly increasing the number of contributions from women in any given class in the Centralized Inquiry group would certainly seem possible. This significant relationship between participation and gender inequities may partially explain why some inquiry-oriented classes were more equitable than others: because they provided more equitable opportunities to participate by gender. Nonetheless, our study design means that we cannot account for how phenomena such as implicit bias, microaggressions, or cultural narratives may have shaped the participation patterns we observed. Future studies that use an experimental design would be better positioned to study if the use of interventions designed to create more equitable levels of classroom participation would result in more equitable student performance, and what variables mediate that relationship.

To further interpret our results, we split our data set into three tertiles. Quantitative analyses showed that the nature of participation in these groups differed significantly. The Centralized Inquiry classes (with large gendered performance differences in favor of men) had the fewest coded student contributions, which was indicative of more instructor-driven discourse—namely, the instructor going back and forth with individual students to elicit their thinking. In contrast, in the Shared Inquiry classes (with a slight gendered performance difference in favor of women) women's participation rates were significantly higher, and more Not Called On contributions were also present. These quantitative differences highlight the overall presence of more discourse in these classes, and more student-to-student interactions in the whole-class discussions.

The other group of classes, Distributed Inquiry, had a slight gendered performance difference in favor of men. What is notable is that these classes actually had the highest relative number of contributions from women—that is, they contributed more than men relative to their demographic representation. However, because of the larger number of contributions overall, women in the Shared Inquiry classes still had the highest participation rates. This finding suggests that the relationship between participation and performance is not a zero-sum game. In the Shared Inquiry classes, both men and women participated more, which resulted in a higher participation rate for women.

Last, we contrast our findings with those of Laursen et al. (2014). Why did the original inquiry-based learning studies find an improvement in gender equity, whereas we found the opposite? A number of differences between the study by Laursen et al. (2014) and the TIMES study that our data were drawn from are worth highlighting. First, the context of the former study focused on mathematics departments that had inquiry-based learning centers, which meant that the idea of inquiry-based learning was systemic within the departments, and students would have had exposure (either directly or indirectly) to inquiry-based learning outside of the focal classes. This contrasts with our study, in which the inquiry-oriented mathematics classes may have functioned more as anomalies within their institutional settings. Second, the inquiry-based learning study focused on self-reported student perceptions, whereas we were looking at actual participation and performance on a content assessment. Third, the nature of inquiry-based learning and inquiry-oriented instruction could be different enough to have different affordances for equitable learning in undergraduate mathematics. For example, in an inquiry-based learning classroom, if students are given an opportunity in advance to choose which theorem they would like to prove and share publicly, it may offer extra preparation time and different entry points for students that allow more students to be publicly successful.

Our findings show that, in inquiry-oriented classes, whether students talk in class, and in particular, whether women talk, matters (using a racially diverse sample). Our data suggest that, in classes that achieve community standards more aligned with Shared Inquiry, women talk more. More broadly, in inquiry-oriented undergraduate mathematics classes in which women talked more, performance was more equitable between men and women. As the field continues to grapple with issues of equity and active learning, sorting out these issues at the levels of content domain, instructional practices, and institutional context will be important.

Limitations

Our findings also have a few key limitations, largely driven by the nature of our secondary analyses. Because we relied on context clues (e.g., student pronouns) to infer gender, we were limited to researchers' perceptions of gender (which we analyzed using binary categories, potentially rendering invisible the contributions of nonbinary students). Further, we were not able to account for students' membership in other social groups in the sample. Similar analyses focused on intersections between gender and race/ethnicity, for example, would be an important area for future study (Bullock, 2018). In acknowledging this limitation, we recognize that research on gender too often centers White women, thereby erasing the experiences of women of color. Although both women of color and White women participated in our study, our inability to link students in the videos to their self-reported demographic information made it impossible to make claims specifically related to women of color in these classrooms. Future study designs—both our own and those of researchers who build on this work—should use an intersectional lens.

Another limitation is that we were not able to link the participation of individual men and women to their performance. Overall, this limited the granularity with which we could make claims using EQUIP-collected data. Accordingly, our statistical models focused on overall participation rates, but did not fully account for the nature of participation. We were still able to link gendered participation to gendered performance in the aggregate but could not account for sources of variation at the individual level, which weakened our statistical power. The use of hierarchical linear modeling in conjunction with individual-level data should certainly be pursued to further strengthen the warrants for claims in this paper. Nonetheless, other work suggests that discourse patterns (specifically, academically productive talk) in the aggregate may be more consequential for individual student performance than individual talk (O'Connor et al., 2017). This is clearly an important area of research that needs to be explored further. Despite these limitations, our work has important implications for mathematics education.

Implications for Practice

First, we echo others who assert the need to move beyond simple recommendations that active learning is good, and lecture is bad, for all students. On the contrary, our study found that some implementations of active learning can disproportionately benefit men, who are already advantaged in mathematics, thus creating further inequity. As long as researchers and policymakers continue to focus on active learning as a panacea, without seriously considering how it can inadvertently create disparities, inequity in mathematics education will certainly persist. Oppressive cultural narratives, implicit bias, and microaggressions are universal phenomena with which any mathematics classroom in the United States will need to contend (Leyva et al., 2021), and unless they are explicitly addressed, they are likely to wreak havoc and cause further inequity. These phenomena have long existed in lecture-based classrooms (Leyva et al., 2021), and simply moving to active learning without attending to them is unlikely to address them. In sum, the idea that active learning is always good is overly simplistic and therefore problematic.

Second, we reflect on the challenge of promoting equitable instruction. The instructors in this study—a majority of whom were women themselves—were well intentioned, had considerable experience teaching mathematics, completed extensive professional development, and were teaching using research-based inquiry-oriented curricula. How then, did gender inequities still emerge? We argue that instructors inherit a problematic patriarchal status quo of mathematics education by default, and only through explicit work can it be overcome. An underlying assumption of the professional development designed by the TIMES project was that a focus on authentic inquiry would also improve equity, but that turned out to be false. Our findings underscore that providing instructors with professional development with an explicit focus on equity is critical. Such professional development must move beyond the theoretical to demonstrate practical instructional moves that teachers can use to combat existing inequities, which instructors must first recognize and then address. In other work, we have demonstrated the potential of providing instructors with data describing patterns of inequities in classroom participation (much like the analytics used in this article), so that they can reflect on their practice and intentionally change their teaching strategies to disrupt those same patterns of inequity (Reinholz, Stone-Johnstone, & Shah, 2020; Reinholz, Stone-Johnstone, White, et al., 2020). Unless instructors receive sustained, ongoing professional development of this type, we imagine that most instructors will find it difficult to address the surmountable inequities in mathematics education.

Third, this work draws attention to the way that both environments and interactions within the environment can contribute to an inequitable gendered learning environment. Consider the Shared Inquiry classes, in which we found significantly higher rates of women's participation and overall Not Called On contributions. We suspect (and have limited evidence) that the instructors in these classes established productive classroom environments that supported their students to build on one another's ideas. Such environments send messages about what it means to do mathematics, and who can do mathematics, by challenging the normative ultra-rational and competitive masculinized notion of mathematics (P. Ernest, 1993). However, because we observed only a single unit well into the semester, we can only speculate about how instructors did this. Nonetheless, we caution instructors from simply enabling Not Called On solicitations and assuming that women will respond. In fact, prior work suggests that men would be most likely to dominate discussion under those circumstances (J. B. Ernest et al., 2019). Future work must attend closely to productive mechanisms for creating gender-equitable inquiry classrooms, because our work shows they do not simply happen by accident.

Conclusion

In contrast to conventional wisdom, we found that active learning does not necessarily lead to improved gender equity in undergraduate mathematics. Rather, across a sample of 20 inquiry-oriented classes, we documented a gendered performance difference between women and men that did not exist in the noninquiry sample. Using a weighted regression analysis, we established a significant link between women's participation rates and gendered performance differences. This empirical link makes an important contribution to the field because it allows us to move beyond student self-reports and theoretical arguments in favor of active learning to track actual levels of student participation. We find that the most consequential factor for improving gender equity in undergraduate mathematics is not simply instructors implementing active learning, but how they implement it.

References

Abdou, C. M., & Fingerhut, A. W. (2014). Stereotype threat among Black and White women in health care settings. *Cultural Diversity and Ethnic Minority Psychology*, *20*(3), 316–323. https://doi.org/10.1037/a0036946

Bando, R., Näslund-Hadley, E., & Gertler, P. (2019). *Effect of inquiry and problem based pedagogy on learning: Evidence from 10 field experiments in four countries* (Working Paper No. 26280). National Bureau of Economic Research. https://doi.org/10.3386/w26280

Banes, L. C., Restani, R. M., Ambrose, R. C., Martin, H. A., & Bayley, R. (2020). Relating performance on written assessments to features of mathematics discussion. *International Journal of Science and Mathematics Education*, *18*(7), 1375–1398. https://doi.org/10.1007/s10763-019-10029-w

Bressoud, D. M. (2021). The strange role of calculus in the United States. *ZDM*, *53*(3), 521–533. https://doi.org/10.1007/s11858-020-01188-0

Bullock, E. C. (2018). Intersectional analysis in critical mathematics education research: A response to figure hiding. *Review of Research in Education*, *42*(1), 122–145. https://doi.org/10.3102/0091732X18759039

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.

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

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*(6), 1045–1060. https://doi.org/10.1037/a0016239

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3

Cooper, K. M., & Brownell, S. E. (2016). Coming out in class: Challenges and benefits of active learning in a biology classroom for LGBTQIA students. *CBE—Life Sciences Education*, *15*(3), Article 37. https://doi.org/10.1187/cbe.16-01-0074

Crenshaw, K. (1990). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

Dasgupta, N., Scircle, M. M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(16), 4988–4993. https://doi.org/10.1073/pnas.1422822112

de Onís, C. M. (2017). What's in an "x"?: An exchange about the politics of "Latinx". *Chiricú Journal*, *1*(2), 78–91. https://doi.org/10.2979/chiricu.1.2.07

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLOS ONE*, *11*(7), Article e0157447. https://doi.org/10.1371/journal.pone.0157447

Engle, R. A. (2012). The Productive Disciplinary Engagement framework: Origins, key concepts, and developments. In D. Yun Dai (Ed.), *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning* (pp. 161–200). Routledge.

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

Ernest, P. (1993). The popular image of mathematics. *Humanistic Mathematics Network Journal*, *1*(8), 53–55. https://doi.org/10.5642/hmnj.199301.08.18

Fortune, N., & Keene, K. A. (2021). Participating in an online working group and reforming instruction: The case of Dr. DM. *International Journal of Research in Undergraduate Mathematics Education*, *7*(1), 107–139. https://doi.org/10.1007/s40753-020-00126-5

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Kluwer.

Garcia, G. A. (2017). Defined by outcomes or culture? Constructing an organizational identity for Hispanic-serving institutions. *American Educational Research Journal*, *54*(1 Suppl.), 111S–134S. https://doi.org/10.3102/0002831216669779

Gholson, M. L. (2016). Clean corners and algebra: A critical examination of the constructed invisibility of Black girls and women in mathematics. *Journal of Negro Education*, *85*(3), 290–301. https://doi.org/10.7709/jnegroeducation.85.3.0290

Gutiérrez, 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. https://doi.org/10.1207/S15327833MTL04023_4

Hall, J., Robinson, T., Flegg, J., & Wilkinson, J. (2020). First-year and final-year undergraduate students' perceptions of university mathematics departments. *Mathematics Education Research Journal*. Advance online publication. https://doi.org/10.1007/s13394-020-00340-z

Hall, W., Keene, K., & Fortune, N. (2016). Measuring student conceptual understanding: The case of Euler's method. In T. Fukawa-Connelly, N. Engelke Infante, M. Wawro, & S. Brown (Eds.), *Proceedings of the 19th annual Conference on Research in Undergraduate Mathematics Education* (pp. 837–842). Special Interest Group of the Mathematics Association of America for Research in Undergraduate Mathematics Education.

Hufferd-Ackles, K., Fuson, K. C., & Sherin, M. G. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education*, *35*(2), 81–116. https://doi.org/10.2307/30034933

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). The median split: Robust, refined, and revived. *Journal of Consumer Psychology*, *25*(4), 690–704. https://doi.org/10.1016/j.jcps.2015.06.014

Johnson, E. (2013). Teachers' mathematical activity in inquiry-oriented instruction. *The Journal of Mathematical Behavior*, *32*(4), 761–775. https://doi.org/10.1016/j.jmathb.2013.03.002

Johnson, E., Andrews-Larson, C., Keene, K., Melhuish, K., Keller, R., & Fortune, N. (2020). Inquiry and gender inequity in the undergraduate mathematics classroom. *Journal for Research in Mathematics Education*, *51*(4), 504–516. https://doi.org/10.5951/jresematheduc-2020-0043

Joseph, N. M. (2017). Math girls: The invisibility of black girls in mathematics. *Virginia Mathematics Teacher*, *44*(1), 46–52. http://www.vctm.org/The-Invisibility-of-Black-Girls-in-Mathematics

Joseph, N. M., Hailu, M., & Boston, D. (2017). Black women's and girls' persistence in the P–20 mathematics pipeline: Two decades of children, youth, and adult education research. *Review of Research in Education*, *41*(1), 203–227. https://doi.org/10.3102/0091732X16689045

Kuster, G., Johnson, E., Keene, K., & Andrews-Larson, C. (2018). Inquiry-oriented instruction: A conceptualization of the instructional principles. *PRIMUS*, *28*(1), 13–30. https://doi.org/10.1080/10511970.2017.1338807

Kuster, G., Johnson, E., Rupnow, R., & Wilhelm, A. G. (2019). The Inquiry-Oriented Instructional Measure. *International Journal of Research in Undergraduate Mathematics Education*, *5*(2), 183–204. https://doi.org/10.1007/s40753-019-00089-2

Lambert, R., & Tan, P. (2017). Conceptualizations of students with and without disabilities as mathematical problem solvers in educational research: A critical review. *Education Sciences*, *7*(2), Article 51. https://doi.org/10.3390/educsci7020051

Langer-Osuna, J. M., & Esmonde, I. (2017). Identity in research on mathematics education. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 637–648). National Council of Teachers of Mathematics.

Larnell, G. V., Boston, D., & Bragelman, J. (2014). The stuff of stereotypes: Toward unpacking identity threats amid African American students' learning experiences. *Journal of Education*, *194*(1), 49–57. https://doi.org/10.1177/002205741419400107

Larsen, S., Johnson, E., & Weber, K. (Eds.). (2013). The Teaching Abstract Algebra for Understanding project: Designing and scaling up a curriculum innovation [Special issue]. *Journal of Mathematical Behavior*, *32*(4).

Laursen, S. L., Hassi, M.-L., Kogan, M., & Weston, T. J. (2014). Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study. *Journal for Research in Mathematics Education*, *45*(4), 406–418. https://doi.org/10.5951/jresematheduc.45.4.0406

Levitt, H. M., Puckett, J. A., Ippolito, M. R., & Horne, S. G. (2012). Sexual minority women's gender identity and expression: Challenges and supports. *Journal of Lesbian Studies*, *16*(2), 153–176. https://doi.org/10.1080/10894160.2011.605009

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

Leyva, L. A., McNeill, R. T., Marshall, B. L., & Guzmán, O. A. (2021). "It seems like they purposefully try to make as many kids drop": An analysis of logics and mechanisms of racial-gendered inequality in introductory mathematics instruction. *The Journal of Higher Education*, *92*(5), 784–814. https://doi.org/10.1080/00221546.2021.1879586

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.

Mathematical Association of America. (2018). *Instructional practices guide*. https://www.dropbox.com/s/42oiptp46i0g2w2/MAA_IP_Guide_V1-2.pdf

McAfee, M. (2014). The kinesiology of race. *Harvard Educational Review*, *84*(4), 468–491. https://doi.org/10.17763/haer.84.4.u3ug18060x847412

McGee, E. O., & Bentley, L. (2017). The troubled success of Black women in STEM. *Cognition and Instruction*, *35*(4), 265–289. https://doi.org/10.1080/07370008.2017.1355211

Melhuish, K. M. (2015). *The design and validation of a group theory concept inventory* [Doctoral dissertation, Portland State University]. PDX Scholar. https://pdxscholar.library.pdx.edu/open_access_etds/2490/

Melhuish, K. (2019). The Group Theory Concept Assessment: A tool for measuring conceptual understanding in introductory group theory. *International Journal of Research in Undergraduate Mathematics Education*, *5*(3), 359–393. https://doi.org/10.1007/s40753-019-00093-6

Mendick, H. (2006). *Masculinities in mathematics*. Open University Press.

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all*. https://www.nctm.org/PtA/

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

Quinn, K. N., Kelley, M. M., McGill, K. L., Smith, E. M., Whipps, Z., & Holmes, N. G. (2020). Group roles in unstructured labs show inequitable gender divide. *Physical Review Physics Education Research*, *16*(1), Article 010129. https://doi.org/10.1103/PhysRevPhysEducRes.16.010129

Rasmussen, C., Keene, K. A., Dunmyre, J., & Fortune, N. (2018). *Inquiry oriented differential equations: Course materials*. https://iode.wordpress.ncsu.edu

Rasmussen, C., & Kwon, O. N. (2007). An inquiry-oriented approach to undergraduate mathematics. *The Journal of Mathematical Behavior*, *26*(3), 189–194. https://doi.org/10.1016/j.jmathb.2007.10.001

Rasmussen, M. L. (2009). Beyond gender identity? *Gender and Education*, *21*(4), 431–447. https://doi.org/10.1080/09540250802473958

Reinholz, D. L. (2018). Reflective apprenticeship for teaching and learning mathematical proof. *Journal of Research in STEM Education*, *4*(1), 68–80. https://doi.org/10.51355/jstem.2018.36

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. https://doi.org/10.5951/jresematheduc.49.2.0140

Reinholz, D. L., Stone-Johnstone, A., & Shah, N. (2020). Walking the walk: Using classroom analytics to support instructors to address implicit bias in teaching. *International Journal for Academic Development*, *25*(3), 259–272. https://doi.org/10.1080/1360144X.2019.1692211

Reinholz, D. L., Stone-Johnstone, A., White, I., Sianez, L. M., & Shah, N. (2020). A pandemic crash course: Learning to teach equitably in synchronous online classes. *CBE—Life Sciences Education*, *19*(4), Article 60. https://doi.org/10.1187/cbe.20-06-0126

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, *50*(4), 1262–1281. https://doi.org/10.1037/a0035073

Sadker, D. M., Sadker, M. P., & Zittleman, K. R. (2009). *Still failing at fairness: How gender bias cheats girls and boys in school and what we can do about it*. Scribner.

Saxe, K., & Braddy, L. (2015). *A common vision for undergraduate mathematical science programs in 2025*. The Mathematical Association of America. https://www.maa.org/sites/default/files/pdf/CommonVisionFinal.pdf

Setren, E., Greenberg, K., Moore, O., & Yankovich, M. (2019). Effects of flipped classroom instruction: Evidence from a randomized trial. *Education Finance and Policy*, *16*(3), 363–387. https://doi.org/10.1162/edfp_a_00314

Shah, N. (2019). "Asians are good at math" is not a compliment: STEM success as a threat to personhood. *Harvard Educational Review*, *89*(4), 661–686. https://doi.org/10.17763/1943-5045-89.4.661

Shah, N., Herbel-Eisenmann, B., & Reinholz, D. (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* (Vol. *5*, pp. 1974–1981). International Society of the Learning Sciences.

Shah, N., & Lewis, C. M. (2019). Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework. *Cognition and Instruction*, *37*(4), 423–452. https://doi.org/10.1080/07370008.2019.1631825

Solomon, Y., Lawson, D., & Croft, T. (2011). Dealing with "fragile identities": Resistance and refiguring in women mathematics students. *Gender and Education*, *23*(5), 565–583. https://doi.org/10.1080/09540253.2010.512270

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

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Staats, C., Capatosto, K., Tenney, L., & Mamo, S. (2017). *State of the science: Implicit bias review.* Kirwan Institute for the Study of Race and Ethnicity, Ohio State University. http://kirwaninstitute.osu.edu/wp-content/uploads/2017/11/2017-SOTS-final-draft-02.pdf

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. https://doi.org/10.1037/0003-066X.52.6.613

Stinson, D. W. (2008). Negotiating sociocultural discourses: The counter-storytelling of academically (and mathematically) successful African American male students. *American Educational Research Journal*, *45*(4), 975–1010. https://doi.org/10.3102/0002831208319723

Suárez-Orozco, C., Casanova, S., Martin, M., Katsiaficas, D., Cuellar, V., Smith, N. A., & Dias, S. I. (2015). Toxic rain in class: Classroom interpersonal microaggressions. *Educational Researcher*, *44*(3), 151–160. https://doi.org/10.3102/0013189X15580314

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., II, Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., . . . Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, *20*(9), 1132–1139. https://doi.org/10.1111/j.1467-9280.2009.02417.x

Wawro, M., Zandieh, M., Rasmussen, C., & Andrews-Larson, C. (2013). *Inquiry oriented linear algebra: Course materials*. http://iola.math.vt.edu/

Wu, S. Y., & Battey, D. (2021). The cultural production of racial narratives about Asian Americans in mathematics. *Journal for Research in Mathematics Education*, *52*(5), 581–614. https://doi.org/10.5951/jresematheduc-2020-0122

Yogeeswaran, K., Devos, T., & Nash, K. (2016). Understanding the nature, measurement, and utility of implicit intergroup biases. In C. G. Sibley & F. K. Barlow (Eds.), *The Cambridge handbook of the psychology of prejudice* (pp. 241–266). Cambridge University Press. https://doi.org/10.1017/9781316161579.011

Authors

Daniel Reinholz, Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182; daniel.reinholz@sdsu.edu

Estrella Johnson, Department of Mathematics, Virginia Tech, Blacksburg, VA 24061; strej@vt.edu

Christine Andrews-Larson, College of Education, Florida State University, Tallahassee, FL 32306; cjlarson@fsu.edu

Amelia Stone-Johnstone, Department of Mathematics, California State University–Fullerton, Fullerton, CA 92831; astonejohnstone@fullerton.edu

Jessica Smith, College of Education, Florida State University, Tallahassee, FL 32306; jls17r@my.fsu.edu

Brooke Mullins, Department of Mathematics & Computer Science; The University of Virginia's College at Wise, Wise, VA 24293; sbs3v@uvawise.edu

Nicholas Fortune, Department of Mathematics, Western Kentucky University, Bowling Green, KY 42101; nicholas.fortune@wku.edu

Karen Keene, Department of STEM Education, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114; Karen.Keene@erau.edu

Niral Shah, Department of Learning Sciences & Human Development, University of Washington–Seattle, Seattle, WA 98105; niral@uw.edu

Submitted February 8, 2021
Accepted May 4, 2021
doi:10.5951/jresematheduc-2020-0143

APPENDIX A

Intersectional Student Demographics From a Subset of the Inquiry-Oriented Instruction Classes

Race/ethnicity	Gender								All	
	Women		Men		Nonbinary		No answer			
White	AA	19	AA	24	AA	0	AA	1	AA	44
	DE	39	DE	66	DE	1	DE	0	DE	106
	Both	58	Both	90	Both	1	Both	1	Both	150
Latinx	AA	1	AA	0	AA	0	AA	0	AA	1
	DE	12	DE	20	DE	0	DE	0	DE	32
	Both	13	Both	20	Both	0	Both	0	Both	32
Asian	AA	1	AA	4	AA	0	AA	0	AA	5
	DE	7	DE	29	DE	0	DE	0	DE	36
	Both	8	Both	33	Both	0	Both	0	Both	41
Black	AA	0	AA	0	AA	0	AA	0	AA	0
	DE	4	DE	14	DE	0	DE	0	DE	18
	Both	4	Both	14	Both	0	Both	0	Both	18
American Indian or Alaskan Native	AA	0	AA	0	AA	0	AA	0	AA	0
	DE	0	DE	0	DE	1	DE	0	DE	1
	Both	0	Both	0	Both	1	Both	0	Both	1
Native Hawaiian or other Pacific Islander	AA	0	AA	0	AA	0	AA	0	AA	0
	DE	0	DE	1	DE	0	DE	0	DE	1
	Both	0	Both	1	Both	0	Both	0	Both	1
Other or bi- or multiracial	AA	2	AA	2	AA	0	AA	0	AA	4
	DE	5	DE	17	DE	1	DE	0	DE	23
	Both	7	Both	19	Both	1	Both	0	Both	27
No race indicated	AA	3	AA	1	AA	0	AA	1	AA	5
	DE	0	DE	3	DE	0	DE	2	DE	5
	Both	3	Both	4	Both	0	Both	3	Both	10
All	AA	26	AA	31	AA	0	AA	2	AA	59
	DE	67	DE	150	DE	3	DE	2	DE	222
	Both	93	Both	181	Both	3	Both	4	Both	281

Note. AA: abstract algebra; DE: differential equations. These data include only students for whom intersectional demographic data were available. Of the nine inquiry-oriented abstract algebra classes included in this study, student race data were collected in only three. Thus, the 281 students in this table represent a subset of the 369 students in the larger sample.

APPENDIX B

Distribution of Student Performance According to Gender

Class	Gender			Average *z*-scores		Gendered performance differences[a]		
	Men	Women	Other	z_Men	z_Women	W – M	W – M (rel.)	*SEM*[b]
				Noninquiry				
DE	74	33	0	−0.35	−0.14	0.21	–	0.09
AA	192	178	4	0.03	−0.13	−0.17	–	0.05
				Group 1				
AA6	15	15	1	0.75	−0.77	−1.51	−1.35	0.22
DE3	9	14	0	0.56	−0.36	−0.92	−1.13	0.25
DE6	26	3	1	−0.51	−1.22	−0.71	−0.92	0.18
DE12	14	2	1	0.21	−0.22	−0.43	−0.64	0.28
AA1	3	8	0	0.00	−0.75	−0.75	−0.58	0.25
DE10	17	2	1	0.34	0.03	−0.31	−0.52	0.28
				Group 2				
DE9	10	6	0	0.3	0.12	−0.19	−0.4	0.21
DE1	15	7	1	−0.1	−0.24	−0.15	−0.36	0.19
AA7	11	8	1	0.1	−0.37	−0.47	−0.30	0.19
AA2	12	7	1	1.19	0.76	−0.43	−0.26	0.27
DE5	7	8	0	0.73	0.68	−0.04	−0.25	0.23
AA5	2	2	0	−0.62	−0.96	−0.34	−0.17	0.50
AA10	5	3	0	0.97	0.68	−0.29	−0.13	0.26
AA4	7	9	0	−0.21	−0.49	−0.28	−0.11	0.24
				Group 3				
DE4	16	6	1	0.11	0.25	0.14	−0.07	0.23
DE11	8	10	0	−0.27	−0.02	0.25	0.04	0.20
AA12	5	2	0	0.02	0.23	0.2	0.37	0.35
DE8	14	5	0	0.41	1.02	0.61	0.40	0.18
AA3	3	2	0	0.57	0.91	0.34	0.51	0.30
DE7	14	4	0	0.03	0.78	0.75	0.54	0.19

Note. AA: abstract algebra; DE: differential equations. We dropped Classrooms DE2, AA8, AA9, and AA11 from our analyses.

[a] Throughout the article we use the (relative) gendered performance difference, but simply use the term "gendered performance difference" for simplicity. For each AA instructor, A_i, we calculated (average A_iWomen *z*-score – average A_iMen *z*-score) – (average noninquiry abstract algebra women *z*-score – average noninquiry abstract algebra men *z*-score). For each DE instructor, D_j, we calculated (average D_jWomen *z*-score – average D_jMen *z*-score) – (average noninquiry differential equations women *z*-score – average noninquiry differential equations men *z*-score).

[b] Standard error of the mean.

APPENDIX C

EQUIP Codes Used to Capture Inquiry-Oriented Discourse

Code	Subcode	Definition
Solicitation method (how is speaker selected)	Group	Instructor calls on a group and a particular student speaks
	Individual	Instructor calls on a student by name
	Volunteer	Instructor calls on a student volunteering to talk
	Random	Instructor uses randomization to identify a speaker
	Not Called On	A student interjects without being called on by instructor
Instructor solicitation (question type)	Why	Instructor asks a student to explain/justify their reasoning
	How	Instructor asks for a student's solution method
	What	Instructor asks a student to read part of a problem, recall a fact, or give a numerical/verbal answer
	Other	Instructor asks a general question (e.g., "What did you think?")
	N/A	Instructor does not ask the student a question
Student talk	Why	Student explains/justifies their reasoning
	How	Student describes a solution method
	What	Student reads part of the problem, recalls a fact, or gives a numerical/verbal answer to a problem
	Other	Student asks a question or says something nonmathematical
Student talk length	21+ words	Student speaks 21+ words consecutively
	5–20 words	Student speaks 5–20 words consecutively
	1–4 words	Student speaks 1–4 words consecutively
Instructor response	Elaborate	Instructor expands on or formalizes the student's idea
	Revoice	Instructor repeats the student's contribution
	Evaluate	Instructor explicitly says the student is correct/incorrect
	Follow-Up	Instructor asks a follow-up question and a new student responds
	N/A	Instructor does not respond to the student's contribution