# Walking the walk: Using classroom analytics to support instructors to address implicit bias in teaching

Daniel L. Reinholz, Amelia Stone-Johnstone

*Department of Mathematics and Statistics, San Diego State University, San Diego, CA, USA*

Niral Shah

*College of Education, University of Washington, Seattle, WA, USA*

Daniel L. Reinholz*
San Diego State University
Department of Mathematics & Statistics
5500 Campanile Drive
San Diego, CA 92182-7720 USA
Phone: 619-594-6191
daniel.reinholz@sdsu.edu
ORCID: 0000-0003-1258-2805
*Corresponding Author

Disclosure Statement: There are no conflicts of interest.

# Walking the walk: Using classroom analytics to support instructors to address implicit bias in teaching

Implicit bias is an issue that all instructors must face in their classrooms. This manuscript describes an academic development method to help instructors address this issue. The method centres on EQUIP (https://www.equip.ninja), a free, web-based application for performing classroom observations that provide data disaggregated by social markers. Instructors engaged in iterative cycles of reflection around EQUIP analytics to track participation in their classes and make measurable changes to their teaching. Here we describe a semester-long learning community in which three university mathematics instructors reflected on EQUIP analytics to reduce implicit bias in their teaching.

Keywords: Gender; implicit bias; mathematics education; race; reflection.

## Introduction

This manuscript focuses on the pervasive problem of implicit bias in schools and universities. Implicit biases are 'discriminatory biases based on implicit attitudes or implicit stereotypes' (Greenwald & Krieger, 2006, p. 951), which subtly shape one's thoughts and actions. How can instructors learn to recognize and address these biases?

We present a low-cost, scalable tool for reducing implicit bias in teaching. EQUIP (**E**quity **Qu**antified **I**n **P**articipation) is a valid and reliable web-based classroom observation tool (https://www.equip.ninja) that provides disaggregated information on the distribution of participation and participation opportunities to particular subgroups (e.g., by race, gender) and individual students (Reinholz, Bradfield, & Apkarian, 2019; Reinholz & Shah, 2018).

This manuscript explores how three university mathematics instructors reflected on and learned to address biases in their teaching. Because racism and sexism are prevalent in mathematics education (Ernest, Reinholz, & Shah, in press; Martin, Rousseau-Anderson, & Shah, 2017), it is an important context for studying bias in

teaching. Through qualitative analyses of participant interviews and learning community meetings, we provide insight into participant learning and the scalability of this approach for academic development. We address two critical research questions:

(1) How do analytics provide useful feedback to instructors to help them recognize racialized and gendered patterns of participation in their classrooms?
(2) What changes do instructors make to their teaching practices upon receiving this feedback?

## Conceptual Frame

### *Using Data to Improve Instruction*

A primary goal of academic development is to create conditions supportive of teaching and learning on a campus (Leibowitz, 2014). Increasingly, data support this goal, for instance, by allowing academic developers to evaluate the impact of their efforts (e.g., Meizlish, Wright, Howard, & Kaplan, 2018). Data also provide insight into classroom teaching and learning environments. For instance, qualitative classroom observations (i.e. pen-and-paper observations) are used to provide feedback to instructors (Wulff & Nyquist, 1986). This simple and customizable approach allows academic developers to provide customized feedback. Technology tools (like the EQUIP app) can streamline observations and automatically generate data analytics.

Data are also a tool to address inequity. For instance, data on racial and gender disparities in student outcomes can spark institutional change (e.g., Bensimon & Malcolm, 2012; Mayer et al., 2014). Similarly, data providing evidence of gender bias in instructor-student interactions can impact teaching practices (Drudy & Chatháin, 2002). Given the promise of these prior studies, we set out to understand the role of data analytics in reducing bias in university teaching.

### *Implicit Bias*

Implicit bias is a form of implicit cognition, which 'influences judgement in a fashion not introspectively known by the actor' (Greenwald & Krieger, 2006, p. 4). Individuals are often unaware of what their biases are—or even that they have biases at all—but their actions are still influenced in subtle ways. Research suggests that biases are reproduced through socialization, beginning at a young age (Yogeeswaran, Devos, & Nash, 2016).

The pernicious impacts of bias are well-documented across domains of public life, including: the workplace, healthcare, criminal justice, and education (Staats, Capatosto, Tenney, & Mamo, 2017). Extensive research shows how bias impacts perceptions of student behaviour, which leads to racialized disparities in discipline (Carter, Skiba, Arredondo, & Pollock, 2017). Bias also impacts perceptions of ability. For example, gender biases cause teachers to systematically underestimate the mathematics proficiency of young girls (Robinson-Cimpian, Lubienski, Ganley, & Copur-Gencturk, 2014). Similarly, STEM instructors tend to underestimate the competence of women in science (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). These biased perceptions mediate instructor expectations for students, which results in disparities in student outcomes (Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010).

Research also documents how bias leads to disparities in participation opportunities based on gender and race (McAfee, 2014; Sadker, Sadker, & Zittleman, 2009; Smith, Andrews-Larson, Reinholz, Stone-Johnstone, & Mullins, 2019). Given the pervasiveness of bias, these findings suggest that such disparities are widespread across university settings, especially in STEM, where racialized and gendered narratives are prevalent (e.g., Leslie, Cimpian, Meyer, & Freeland, 2015; Shah, 2017). Because

participation is key to learning (Michaels, O'Connor, Hall, & Resnick, 2010), this implies some students receive fewer, or lower quality opportunities to learn.

Biases also often lead to microaggressions, which are 'brief, everyday exchanges that send denigrating messages to people [who are minoritized] because they belong to a [minoritized] group' (Sue et al., 2007, p. 272). Microaggressions typically manifest through brief exchanges, delivered through dismissive looks, gestures, tones, or comments. Microaggressions have been likened to a 'death by a thousand cuts' (Nadal et al., 2011), because a high frequency of these subtle events can have a devastating impact. In this way, bias significantly contributes to the 'chilly climate' of STEM, which consists of learning environments that are unwelcoming for women and students of colour (Johnson, 2007; Seymour & Hewitt, 1997).

### *Addressing Bias*

The efficacy and longevity of interventions to address implicit bias is an area of active study (Staats et al., 2017). Nonetheless, these studies suggest that attempting to eliminate one's biases or simply pretending that they do not exist is not particularly effective (DiAngelo, 2018). Biases need to be recognized and addressed. Our approach involves providing *feedback* to an instructor, consisting of data analytics that facilitate reasoning about patterns of classroom participation, coupled with community-based academic development. The data alone do not constitute feedback; they provide the foundation for exploring racialized and gendered patterns of participation.

Despite intentions to engage all students in their classrooms, an instructor may be unaware that they are failing to do so. Thus, data analytics can help an instructor bridge the gap between desired results and actual performance (Black & Wiliam, 2009; Sadler, 1989). To support sensemaking around these data, we follow principles of good

feedback; it is: nonevaluative, supportive, timely, and specific (Shute, 2008). Such feedback aims to directly inform practice.

This feedback supports instructor reflection, which is *the act of processing an experience, action, or practice to inform future action* (Reinholz, 2016). Reflection helps instructors revisit a past experience, re-evaluate the experience, and determine future actions (Boud, Keogh, & Walker, 1996). By reflecting *retrospectively* on data analytics, instructors can develop new ways of thinking to guide their future moment-to-moment decisions (Reinholz, 2016). In this way, analytics help slow down the automatic associations of bias, so that an instructor can become aware of them, interrogate them, and change future actions.

## EQUIP

### *The Tool*

EQUIP is a free, fully-customizable web app (https://equip.ninja) that automatically generates analytics and streamlines the coding of classroom video (Reinholz & Shah, 2018). It does not evaluate instructors or tell them how to teach; instead, it provides data to support conversations about equitable teaching. In the present study, participants received data from the report functionality of EQUIP (see Appendix A). An EQUIP report has three parts: (1) and overall summary of classroom participation, (2) a summary of individual student participation, and (3) a summary of group participation by each dimension. EQUIP also includes a variety of interactive analytics (e.g., histograms, heatmaps, and timeline graphs) that were developed after this study was conducted. These analytics are designed to process the raw data from a classroom observation and present it in a format that is conducive to instructor reflection. After these analytics are generated by EQUIP, they are shared with instructors by the academic developer.

EQUIP allows a user to create any number of custom demographic categories and features of talk to code (called discourse dimensions). The seven default discourse dimensions – (1) discourse type, (2) student talk length, (3) student talk type, (4) teacher solicitation method, (5) wait time, (6) teacher solicitation type, and (7) explicit evaluation – were established in prior work from a review of the literature (Reinholz & Shah, 2018); here we used dimensions (2), (3), and (6) to track the types and quality of student participation and participation opportunities.

EQUIP allows a user to code each verbal contribution a student makes, which can be used to document student-instructor and student-student interactions. To use the app, an observer first sets up a classroom by choosing their discourse dimensions, demographic categories, and sets up a student roster that includes demographic information (provided by the cooperating instructor). This creates a seating chart in the app, and each time a student participates a user clicks on the student's desk to code the participation. The app can be used in real-time in the classroom or it can be used to code video.

### *Building a Learning Community*

This study involved a semester-long learning community organized around reflection on analytics. To begin, each participant met individually with the academic developer to discuss their equity goals. Participants chose the demographics (e.g., race, gender, hours worked) and discourse dimensions (e.g., length of talk, type of talk) they wanted to focus on. After the interviews, each participant shared seating charts and student demographics to support observations. The instructors determined student demographics based on how they identified their students, with support from other sources (e.g., student surveys). We focus on instructor identification, because it is most closely related to each instructor's own biases.

The learning community was organized around a series of four reflection cycles. Each cycle involved: (1) recording each instructor's teaching, (2) coding videos and generating analytics, and (3) debriefing. During debriefs participants set specific action plans to change teaching practices, which could be measured in the next cycle. The second author coded videos and provided analytic reports to participants. These reports were discussed during debriefs, and at times they also initiated email conversations about equity. Reports were provided to describe individual observations, and aggregate reports were given that described all instruction to date. The first two authors were the academic developers who organized the meetings, designed activities, and ensured that discussions remained productive.

### *Principles for Facilitation*

Facilitation followed the four features of good feedback. To provide *non-evaluative* feedback, participants were provided with disaggregated data analytics, and set their own equity goals. To mitigate judgment, academic developers acknowledged the systemic and societal foundations of bias, and recognized that instructors are teaching in a flawed system. The developers were supportive, caring, and positive when interacting with participants, and worked to foster relationships between participants to support the vulnerability, trust, safety, and respect required to work on issues of racial and gender bias.

To provide *supportive* feedback, data were reflected upon in a group setting. Participants developed camaraderie as they collectively worked towards the larger goal of more equitable teaching. The developers built on existing relationships between participants, and the lead author attended a community gathering with participants.

Feedback provided was *specific*. Because data focused on race and gender explicitly, conversations were shifted away from discussing equity in general terms. This supported instructors to consider their impact on specific students and groups of students. Because instructors could track participation for individual students in their classes, they could design, implement, and measure targeted changes to their practice.

Finally, feedback was *timely,* because instructors received actionable data they could use to adjust teaching practices. Through multiple rounds of reflection, participants could see the impact of changes to their teaching. The whole process was designed to be formative. No single observation could capture everything about the instructor's teaching, but the goal was to move towards more equitable participation over time.

## Method

### *Participants and context*

Participants were three mathematics instructors working at Minority-Serving Institutions in the USA. Thus, the student populations involved were more diverse along a number of dimensions (e.g., race, first-generation status, socioeconomic status) than one would encounter at Primarily White Institutions. The participants were recruited from a professional organization that focuses on equity in mathematics. Thus, they all had commitments to racial and gender justice, but had not necessarily received academic development on how to enact this commitment through their teaching. In addition, one participant had existing, but separate, professional relationships with the other two participants. This helped increase trust.

The instructors were teaching in relatively small classrooms, with fewer than 40 students in each. They used a variety of teaching methods, including: lecture, whole-

class discussion, and small group work. While EQUIP can be used to study participation across different classroom formats (e.g., small group, whole class), in this study, we focused only on whole-class discussions. We chose to focus on only one aspect of practice because it was less resource intensive, and thus would be a practical method for use in many academic development units (e.g., by a Centre for Teaching and Learning on a campus).

***Data collection***

Pre- and post-interviews were conducted with each participant. Interviews focused on: teaching experience, conceptions of equity, specific equitable teaching practices, interpretation of EQUIP analytics, and experiences in the learning community. In addition, all meetings were recorded and transcribed for analysis. The end-of-semester interview protocol is provided in Appendix A.

Prior to the start of the study, instructors discussed the goals of the learning community with the students in their classrooms, and students were given the opportunity to opt out of the study. During each reflection cycle, participants submitted videos of their teaching (1-2 hours in length). Videos were coded using the EQUIP web app, and analytic reports were sent to the participants (see Appendix B for a sample report). These analytics provided context for understanding the learning community meetings, and are also a source of data for tracking changes in teaching.

Due to logistical constraints and other unforeseen circumstances (e.g., natural disasters), not all instructors collected data for each reflection cycle. A summary of collected data and participant demographics are given in Table 1.

<INSERT Table 1>

### *Analytic methods*

The learning community meetings and pre- and post-interviews were the primary data sources for analysis. Data analysis was guided by the four criteria for good feedback (Shute, 2008) as a conceptual framework for interpreting participants' responses. The analysis followed a quasi-deductive approach (Miles & Huberman, 1994). The first author began by reading through all transcripts multiple times and tagging instances where the participants appeared to invoke the principles of feedback. This tagging process was followed-up by a second round of coding that involved categorizing the tagged segments in the four categories. We also supplemented this analysis with quantitative metrics describing changes in instructor practices.

Below, we aim to tell the stories of the participants using their own language. For this reason, we describe each of the themes by using quotes from all three of the instructors. By organizing our analyses around an external framework (i.e. criteria for good feedback), we aimed to speak back to the existing literature and create generalizable knowledge. Upon completing this analysis and connecting the themes, we returned to the dataset to search for other instances of instructor transcript that agreed or disagreed with our interpretations, and updated the narrative accordingly.

## Results

The results section is organized into four parts, related to the four categories of good feedback practice. In each of these subsections, we provide a general discussion of our findings as well as illustrative quotes from each instructor.

### *Non-Evaluative Feedback: Developing a New Lens for Seeing Teaching*

Data analytics provided instructors with a new lens to view their classrooms. By design, this lens would be non-evaluative, as instructors would set their own instructional goals. All participants described how data helped them think differently about their classrooms. For example, Sherry noted that the data 'provided a lens I wouldn't otherwise have or be able to measure.' Similarly, Kevin emphasized one of the greatest strengths of the data was,

> Just bringing that awareness. I mean we don't have a mirror in our classrooms. We're focused on what we're trying to get through. We're trying to get through that material.

Kevin emphasized how the data helped him think differently about his teaching, even though it did not prescribe a specific outcome for him to achieve. Brian also mentioned how the data helped him move beyond talking 'about equity all day,' to instead 're-evaluate' his own practice and think about areas for improvement,

> Thinking about equity and actually doing the hard work to make changes within yourself and want to actually improve equitable outcomes, they're just two different things. But this helps me put my finger on that a little bit more. Like we can talk about equity all day, but it doesn't mean- it's a different level to be able to really want to re-evaluate your own practice and step down from that pedestal that as faculty members we like to sit on and realize that there might be not just some things you could be doing different, but there's a lot of room for growth for everyone, no matter where you are in your practice.

These quotes illustrate how data provided participants with a new way of looking at their classrooms.

### *Supportive Feedback: Developing a Community of Learners*

Participants found that having a supportive community was a crucial aspect of their learning experience. Brian described this as follows,

> To me, having a group of people to analyze this data with and share back ideas on how to talk about what it was saying and give concrete ideas on how to address these issues was really helpful.

While the data were valuable, Brian noted that the community was what helped him actually change his practice. All participants were clearly invested in the community, and actively provided supportive feedback to each other.

Sherry also valued the 'community of learners,' but acknowledged it could be 'awkward' to talk about racial dynamics in her teaching. Sherry emphasized how Brian's humility in reflecting on his practice helped her 'feel safe,' and realize she did not 'have to be perfect either.' Sherry's comments emphasize the need to build trust, which was supported by her prior relationship with Brian and his humility. Sherry also noted that it was 'easier' to talk about gender than race, which may reflect the current climate in the US.

Kevin agreed that the community was 'a good group,' and that he was 'comfortable' with Brian and the academic developers. He also emphasized the need to build 'community safety,' because he had no prior relationship with Sherry. Kevin was cautious as a person of colour, having had negative experiences in other groups focused on race. For example, he described how when people espouse colour-blind statements such as 'I treat everybody the same,' they make the environment unsafe for people of colour. Upon reflection, Kevin described how the learning community put him in his 'students' shoes,' and made him realize that are also times he does not 'feel safe,' and would 'rather just stay quiet.' Thus, being in the vulnerable position of sharing his classroom teaching helped him better empathize with his students.

### *Specific Feedback: Moving Beyond Equity in General*

Participants were provided with disaggregated data analytics to help them focus on specific students or groups of students, rather than a generalized notion of equity. Brian articulated how specific data forced him to move away from post-hoc storytelling,

> Having data about what is happening is way more useful than I ever imagined. Because without it you can tell whatever story you want, right, about what happened in the classroom. You can focus on whatever part of that day or whatever student interactions were so positive that you want to take away, and that can be your whole premise for measuring your effectiveness when it comes to equity issues in the classroom.

Brian articulated the role of confirmation bias, which makes it easy to focus on one or two events that made him feel like an equitable teacher. However, with actual data that spanned the whole class session, it helped him see the whole picture. Brian continued to say that 'data that can tell you a story about your class was really eye opening in a way that I didn't expect.' Still, Brian concluded that it was 'challenging' to 'really consider your own biases and practices' and address them on top of all of the demands of teaching mathematics. Similarly, Sherry noted it was 'difficult to change in one semester' and Kevin stated how changes for him 'were slow.'

Reflecting on analytics also made Sherry more aware of her own biases. In particular, she talked about her tendency to gravitate towards calling on men in her class,

> As we discussed, it started out pretty much all male in terms of the data. And I think that's my implicit bias as well, is that I gravitate towards calling on males in math and just the general climate. I think in terms of female role models and stuff like that, I'm older than most of the female undergrads can relate to so it's not that helpful anymore.

What Sherry described was a specific bias related to men in her classroom. As described in the next section, this provided her with an actionable target to improve equity in her teaching.

Finally, as described above, Kevin likened the data to a 'mirror' in his classroom. He further elaborated that this could help bring awareness to 'who's answering, gender-wise' and 'race-wise' in the classroom. Thus, the data were not just about improving equity generally, but helping participants be very concrete about what they were looking at with regards to equity, and how it might impact their teaching practices.

***Timely Feedback: Changing Teaching Practices***

Participants received timely feedback, which they could use to change their practice. For example, Brian described a number of concrete changes he made in response to the analytics, including regrouping students, calling on particular students, and 'probing deeper into student responses.' to get more students involved in his class. These changes could be seen in the data (see Figure 1). With the exception of Round 3, there was a general upward trend for more students participating during each lesson in Brian's teaching. A similar trend was found for Sherry, based on her strategies of changing the seating charts or intentionally calling on certain sections of the room.

<INSERT FIGURE 1>

Sherry also noted that she 'became more conscious' of her questions, and she attempted to incorporate more 'why' questions (asking for explanations) as a follow-up to what questions (asking for answers). Figure 2 shows the percentage of high-level questions (how and why) during any given lesson. The figure shows, that for both Brian and Sherry, after receiving their first round of analytic feedback they both increased

high-level questions in their teaching, which was one of their stated goals. However, this level of questioning was not maintained into the later rounds of teaching.

For instance, during Brian's third observation, he asked many 'other' questions, such as 'what do you think about that…?' These questions tended to result in 'what' responses from students, which may have been the result of norms already set earlier in the semester. Given that norms are established early in the semester, there may be some limitations to the changes that can be seen in a single semester.

<INSERT FIGURE 2>

Upon noticing her tendency to call on more men in her class, Sherry intentionally included more women in the discussion (see Figure 3). This was a notable improvement of participation by women in her class, who comprised 36% of the class.

<INSERT FIGURE 3>

Not all changes could be seen in the data. For example, Brian began 'paying a lot of attention' to the three Filipino students in his class after the first round of observations showed they were not participating. But this did not 'show up in the data.' It could be that Brian mostly gave this attention during small group work.

Kevin only had one round of analytics, so we could not track changes in his teaching. Still he stated a goal of 'injecting how and why questions,' and that his changes would be in terms of what he was 'doing' and how he was 'viewing' his class.

**Limitations**

Our study had a number of limitations, given the small number of participating instructors, the relatively small classroom sizes they were teaching in, and our emphasis only on whole-class discussions. In addition, we focused primarily on the data analytics, without triangulating other data sources (e.g., student surveys) that could provide deeper feedback to an instructor. These are all issues to be addressed for future work. Moreover, the instructors in this study identified the demographic information of their students, and it would be valuable to compare instructor identification and student self-identification. Finally, another research goal is to perform a larger study that involves a comparison group, so that it will be possible to measure the impact of the professional development on student outcomes. Such a study could also include measures of instructors' biases, to see if bias itself is actually reduced, in addition to mitigating the effects of bias.

**Discussion**

The primary goals of our study were to understand how analytics could help instructors reflect on implicit bias in their teaching, and whether this was a productive approach for academic developers to use. Related to our first research question, the analytics helped participants recognize racialized and gendered patterns in their classrooms. Addressing our second research question, we found that instructors made a number of concrete changes to their teaching practices. These changes focused on the types of questions that were asked of students, how student seating was arranged in the room, and which students were chosen to participate during discussion. Thus, this study provides evidence that technology can help generate data analytics to providing useful formative feedback to instructors (Sadler, 1989, Shute 2008).

The impact of these changes could be seen in the data, as the overall percentage of students participating and the quality of questions asked improved. Although it was not a focus of this study, given the close link between participation and learning (Michaels et al., 2010), we suspect this would lead to more equitable learning outcomes for students. These results generally suggest that reflection on instructional analytics can be a productive way for instructors to engage with their implicit biases and change their teaching practices. Given that EQUIP is a free, customizable tool, we believe this illustrates its promise for use by academic developers in activities including: consultations, workshops, and learning communities.

Our results also highlight the challenges of supporting reflection on bias. It is important to build safety and trust within a learning community, and based on Brian's reflections about community safety, this was evidently only achieved to some extent within the current study. In addition, we found that changes in teaching practices were not always uniform, or could not always be captured by the data. All of the instructors noted that they had greater awareness, but they recognized that change takes time, and it was not always clear to them what to do with their newfound awareness. These findings are notable, given that each of the participants had extensive teaching experience, were working in diverse instructional settings, and had deep commitments to equity, but they still grappled with equity in their classrooms. As such, it is likely that instructors across a wide variety of institutional settings and awareness of equity issues could benefit from this process. Given these challenges, we believe it is important for academic developers to cultivate expertise specifically working around issues of equity and bias, so that they can most effectively support instructors.

Our study also provides insight into a number of productive avenues for future work. For example, Brian's reflection of not knowing what to do with his greater

awareness suggests that academic developers may benefit from developing a suite of equity-focused instructional practices that they could share with members of their learning communities. In addition, the learning community in this study included no explicit instruction about bias, cultural competence, microaggressions, or related phenomena. They were simply addressed as they came up in conversation. In this way, we hypothesize that reflection on analytics supported by an explicit curriculum could be particularly productive.

For academic developers, this work provides concrete strategies for engaging instructors to address implicit biases in their teaching. Moreover, given that racism and sexism are prevalent in mathematics education (Martin et al., 2017), the thoughtful reflections of the three instructors in this study are encouraging. The participants modelled the vulnerability required to address sensitive issues and change their own teaching practices. Still we recognize that these particular instructors already had commitments to equity, which influenced their ability to meaningfully grapple with the data. Even though the study took place with university mathematics instructors, the issues and approach are generalizable to a variety of domains, and can inform academic development practices much more broadly. If these practices were taken up by a permanent campus unit (e.g., a Centre for Teaching and Learning), it would have the potential to increase equity on a campus on an ongoing basis.

**Author Biographies**

Daniel L. Reinholz, is an Assistant Professor of Mathematics Education at San Diego State University. Broadly speaking, Dr. Reinholz's research focuses on creating tools for educational transformation, to improve equity and mitigate systemic oppression. His

research is primarily situated within three interrelated areas: classroom design, racial and gender equity, and systemic change.

Amelia Stone-Johnstone is a doctoral student in the Math and Science Education (MSED) joint doctoral program between SDSU and UCSD. Amelia has been working with Dr. Reinholz to study the use of EQUIP to promote equity in mathematics classrooms.

Niral Shah is an Assistant Professor of Mathematics Education at Michigan State University. His research focuses on equity and implicit bias in STEM education. Although mathematics is often seen as "neutral" and "race-free," Shah's research shows that math classrooms are highly racialized spaces.

**References**

Bensimon, E. S., & Malcolm, L. (Eds.). (2012). *Confronting equity issues on campus: Implementing the equity scorecard in theory and practice*. Sterling, VA: Stylus Publishing.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Boud, D., Keogh, R., & Walker, D. (1996). Promoting reflection in learning: A model. In *Boundaries of Adult Learning* (Vol. 1, pp. 32–56). New York, NY: Routledge.

Carter, P. L., Skiba, R., Arredondo, M. I., & Pollock, M. (2017). You Can't Fix What You Don't Look At: Acknowledging Race in Addressing Racial Discipline

Disparities. *Urban Education*, *52*(2), 207–235. https://doi.org/10.1177/0042085916660350

DiAngelo, R. (2018). *White Fragility: Why It's So Hard for White People to Talk About Racism*. Boston, MA: Beacon Press.

Drudy, S., & Chatháin, M. Ú. (2002). Gender Effects in Classroom Interaction: Data Collection, Self-Analysis and Reflection. *Evaluation & Research in Education*, *16*(1), 34–50. https://doi.org/10.1080/09500790208667005

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

Greenwald, A. G., & Krieger, L. H. (2006). Implicit Bias: Scientific Foundations. *California Law Review*, *94*(4), 945–967. https://doi.org/10.2307/20439056

Johnson, A. C. (2007). Unintended consequences: How science professors discourage women of color. *Science Education*, *91*(5), 805–821. https://doi.org/10.1002/sce.20208

Leibowitz, B. (2014). Reflections on academic development: What is in a name? *International Journal for Academic Development*, *19*(4), 357–360. https://doi.org/10.1080/1360144X.2014.969978

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265. https://doi.org/10.1126/science.1261375

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

Mayer, A. K., Cerna, O., Cullinan, D., Fong, K., Rutschow, E. Z., & Jenkins, D. (2014). *Moving Ahead with Institutional Change: Lessons from the First Round of Achieving the Dream Community Colleges.* New York, NY: MDRC.

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

Meizlish, D. S., Wright, M. C., Howard, J., & Kaplan, M. L. (2018). Measuring the impact of a new faculty program using institutional data. *International Journal for Academic Development*, *23*(2), 72–85.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. B. (2010). *Accountable Talk® Sourcebook*. Pittsburgh, PA: Institute for Learning.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook. Second Edition*. Thousand Oaks: Sage Publications.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109

Nadal, K. L., Issa, M.-A., Leon, J., Meterko, V., Wideman, M., & Wong, Y. (2011). Sexual Orientation Microaggressions: "Death by a Thousand Cuts" for Lesbian, Gay, and Bisexual Youth. *Journal of LGBT Youth*, *8*(3), 234–259. https://doi.org/10.1080/19361653.2011.584204

Reinholz, D. L. (2016). Developing mathematical practices through reflection cycles. *Mathematics Education Research Journal*, *28*(3), 441–455. https://doi.org/10.1007/s13394-016-0175-1

Reinholz, D. L., Bradfield, K., & Apkarian, N. (2019). Using Analytics to Support Instructor Reflection on Student Participation in a Discourse-Focused

Undergraduate Mathematics Classroom. *International Journal of Research in Undergraduate Mathematics Education*. https://doi.org/10.1007/s40753-019-00084-7

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.

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

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences* (Vol. 12). Boulder, CO: Westview Press.

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

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Smith, J., Andrews-Larson, C., Reinholz, D. L., Stone-Johnstone, A., & Mullins, B. (2019). Examined Inquiry-Oriented Instructional Moves with an Eye Toward Gender Equity. *Proceedings of the 2019 Conference on Research in*

*Undergraduate Mathematics Education*. Presented at the Oklahoma City, OK. Oklahoma City, OK.

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

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271.

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

Wulff, D. H., & Nyquist, J. D. (1986). Using Qualitative Methods to Generate Data for Instructional Development. *To Improve the Academy*, *5*(1), 37–46. https://doi.org/10.1002/j.2334-4822.1986.tb00085.x

Yogeeswaran, K., Devos, T., & Nash, K. (2016). Understanding the nature, measurement, and utility of implicit intergroup biases. In C. Sibley & F. Barlow (Eds.), *The Cambridge Handbook of the Psychology of Prejudice* (pp. 241–266). Cambridge: Cambridge University Press.

## Appendix A: Interview Protocol

**General**

1. How was your semester? How did things go with your focal class?
2. Did anything stand out for you in your experience working with EQUIP?

**Overall Classroom Data**

1. What did you notice about your teaching and student participation at the classroom level?
2. How easy was it to make sense of the data? What could have been improved?

3. What changes did you make this semester in your teaching using these data? What changes would you make in the future?

**Individual Student Data**

1. What did you notice about your teaching and student participation in terms of individuals?
2. Did EQUIP draw your attention to any particular students? Tell me more.
3. How easy was it to make sense of the data? What could have been improved?
4. What changes did you make this semester in your teaching using these data? What changes would you make in the future?

**Group-Level Data**

1. Let's talk about race.
    a. What did you notice about your teaching and student participation in terms of racial groups?
        i. What positive trends did you notice in the data?
        ii. What inequities did you notice?
    b. How would you explain these patterns?
        i. Are you surprised or did you expect this?
    c. What kinds of specific things did you do to support students from particular groups? Black students? Latinx students?
    d. How did you feel talking about race and racial equity in the EQUIP groups?
    e. Did you have any insights about particular groups of students, or did EQUIP draw your attention to them in ways you would not have otherwise?
    f. How easy was it to make sense of the data? What could have been improved?
    g. What changes did you make this semester in your teaching using these data? What changes would you make in the future?
2. Repeat for gender, hours worked.

**Professional Development and EQUIP**

1. What did you think about the frequency of our meetings? Would you prefer more, less?
2. Would you participate in this type of professional development again? Would you recommend that others do?
3. What type of support do you think would be most helpful when it comes to working on equity as an instructor?
4. Let's return to EQUIP.
    a. What benefits did you find working with it?
    b. What were some of the cons/drawbacks?
    c. How could the tool be improved?
5. Do you feel like you think differently about equity after doing this work?

**Overall**

1. What are you most proud of this year as it relates to your equitable teaching practices?
2. What equity goals do you have for yourself next year?
3. What did you learn about yourself as a teacher and as a person through this research?
4. What will you take with you to your future teaching?
5. Is there anything else we should know?

Table 1. Participant demographics and data collection

| Name | Gender | Race | Years of Teaching Experience | Number of Observations |
|---|---|---|---|---|
| Sherry | Woman | White | 30 | 3 |
| Kevin | Man | Latinx | 20 | 1 |
| Brian | Man | Black | 15 | 4 |

Figure Captions

Figure 1. Percentage of students participating.

Figure 2. Percentage of high-level questions during a lesson.

Figure 3. Percentage of men and women participating in Sherry's class.