Running head(s):
Equity Analytics
Daniel L. Reinholz and Niral Shah

Equity Analytics: A Methodological Approach for Quantifying Participation Patterns in Mathematics Classroom Discourse

Daniel L. Reinholz
San Diego State University

Niral Shah
Michigan State University

<below rule, on p. 1>

Both authors contributed equally to this article.

We thank Deborah Ball, Victoria Hand, Chuck Munter, and Alan Schoenfeld for their meaningful feedback on earlier versions of this manuscript. We are also grateful to our team of graduate and undergraduate students at Michigan State University who supported the development of EQUIP.

Abstract

Equity in mathematics classroom discourse is a pressing concern, but analyzing issues of equity using observational tools remains a challenge. In this article, we propose equity analytics as a quantitative approach to analyzing aspects of equity and inequity in classrooms. We introduce a classroom observation tool that focuses on relatively low-inference dimensions of classroom discourse, which are cross-tabulated with demographic markers (e.g., gender, race) to identify patterns of more and less equitable participation within and across lessons. We argue that equity analytics can support researchers and practitioners in identifying subtle patterns of inequity in classroom discourse. As we show, even in classrooms with highly experienced, equity-minded teachers, subtle inequities can emerge that are detectable through this quantitative methodology. To conclude, we discuss how equity analytics can complement qualitative approaches in the study of equity and inequity in classrooms.

Key words: Classroom discourse; Equity; Gender; Observational research; Race

Inequity in mathematics education remains a pressing concern (National Council of Teachers of Mathematics, 2014; National Research Council, 2013; President's Council of Advisors on Science and Technology, 2012). Research shows that the persistent underrepresentation of women and non-Asian people of color in mathematics-intensive fields is related to inequitable access to opportunities to learn (Darling-Hammond, 2010; Oakes, 2005). Although the national conversation has focused on representation (e.g., enrollment in advanced mathematics courses and majors), a growing literature has examined equity issues within everyday classroom discourse (Esmonde & Langer-Osuna, 2013; Herbel-Eisenmann, Choppin, Wagner, & Pimm, 2011; Moschkovich, 2011). This research has primarily employed qualitative methodologies to illuminate the nuanced ways in which inequities can emerge in classroom interaction, particularly with respect to gender, race, class, and other social markers. However, relying solely on qualitative methodologies can limit the potential for research to influence policy and practice related to equity.

In the United States alone, there are over 100,000 K–12 schools, which contain over 1,000,000 mathematics classrooms. Across these classrooms, students differ in their educational needs and their available opportunities to learn. Although qualitative methods reveal the nuanced ways in which inequities can emerge in classrooms, the resource-intensive nature of in-depth qualitative work poses an obstacle to its widespread use. What is needed is a simple but powerful methodology for analyzing patterns of equity and inequity in classroom discourse that is adaptable across a diverse array of classroom contexts.

We propose *equity analytics* as a quantitative approach to complement in-depth qualitative studies of equity. Equity analytics involves identifying and analyzing patterns of equity and inequity in classroom discourse. The term *analytics* refers to the use of statistical techniques to analyze patterns in large data sets (i.e., "big data"). Analytics has been used in fields ranging from health care to professional sports. In education, researchers have used analytics to parse complex patterns in student performance (Greller & Drachsler, 2012; Long & Siemens, 2011). However, analytics has not been used to study equity patterns in classroom discourse.

To demonstrate this approach, we introduce a classroom observation tool called EQUIP (Equity QUantified In Participation). EQUIP utilizes equity analytics to illuminate patterns in student participation and the opportunities made available for students to participate in classroom discourse. This contrasts with the few existing observation tools in mathematics education that explicitly attend to issues of equity because they are designed to make holistic claims about the overall quality of mathematical discourse in classrooms (see Boston, 2012; Goffney, 2010; Schoenfeld, 2014). However, these tools were not designed to answer a critical question: *Who* is getting access to opportunities to participate in that mathematical discourse? Without the ability to disaggregate at the level of groups (e.g., by race, gender, or language proficiency) and individual students, we argue that the potential for observation tools to support efforts to make classrooms more equitable is limited. The methodological approach that we describe in this article is intended to address this issue.

The present work is organized around the following research question: What are the affordances and limitations of a quantitative approach to measuring aspects of equity and inequity in classroom discourse? We begin by reviewing prior research on equity issues in classroom discourse. Next, we consider how existing observation tools have operationalized this literature and also discuss some of the shortcomings in those efforts. As a way forward, we detail the theoretical basis for equity analytics, which is grounded in a definition of equity that treats equality as a necessary—albeit insufficient—waypoint towards equity. After laying this conceptual foundation, we use EQUIP to analyze a large data set of teacher and student discourse in an elementary mathematics classroom led by a highly experienced, equity-minded teacher. The findings illustrate how equity analytics can generate actionable insights for researchers and practitioners to address patterns of inequity. We conclude by discussing how equity analytics can productively complement qualitative analyses in the service of more robust, integrated approaches to researching equity in mathematics classrooms. We argue that such mixed-methods approaches should involve triangulating multiple types of data collected from the perspectives of multiple actors in the classroom. Further, we consider how tools like EQUIP might support practitioners in reflecting on equity issues in their own practice, particularly with respect to implicit bias.

Prior Research on Equity and Participation in Mathematics Classroom Discourse

Research supports the relationship between discourse and learning (Ing et al., 2015; Lampert, 1990; National Research Council, 2000; Resnick, Michaels, & O'Connor, 2010; Sfard, 2008). According to Turner, Dominguez, Maldonado, and Empson (2013),

> discussions afford opportunities for students to do such things as explain and justify solution strategies, pose questions, and articulate connections between mathematical ideas. In other words, they afford opportunities for students to take on *agentive problem-solving* roles and to participate in ways that can impact students' dispositions toward the subject and, over time, their sense of themselves as competent, agentive learners. (p. 203)

All students need opportunities to participate in the discursive practices constitutive of the learning process. For the purpose of this article, we focus on *participatory equity*, a particular form of equity defined as a condition in which both participation and participation opportunities are distributed fairly during the learning process (Shah & Lewis, in press).

A number of studies in mathematics education have linked participation in classroom discourse to issues of equity (Bianchini, 1997; Esmonde, 2009; Herbel-Eisenmann et al., 2011; Moschkovich, 2011; Weber, Radu, Mueller, Powell, & Maher, 2010). These scholars have conceptualized "discourse" in a variety of ways (Wagner, Herbel-Eisenmann, & Choppin, 2011). One approach to analyzing the discourse–equity relation has been to consider students' use of nondominant linguistic and cultural practices. Several studies have found that marginalization can occur when students use languages or engage in discursive practices not recognized or valued in learning environments organized by the dominant culture (Moschkovich, 2011; Parks, 2010; Turner, Dominguez, Maldonado, & Empson, 2013). Other studies—including some of our own prior work—have focused on cultural narratives and broader societal discourses (Foucault, 1975; Gee, 2011) and how they position students with identities as mathematics learners (Esmonde & Langer-Osuna, 2013; Langer-Osuna, 2011; McGee & Martin, 2011; Nasir & Shah, 2011; Shah, 2017; Stinson, 2008). From this perspective, classroom discourse is a site where identities are constantly negotiated (Esmonde, 2009; Wood, 2013) and where inequities can emerge in students' access to identities as capable learners of mathematics. In terms of methodology, with few exceptions (e.g., Battey, Neal, Leyva, & Adams-Wiggins, 2016), most classroom-based studies of equity issues have used qualitative methods.

Consistent with much of the literature, we conceptualize discourse in terms of the oral[1] practices shown to be central to learning (e.g., asking questions, explaining one's thinking). We view the distribution of opportunities to participate in classroom discourse—and the extent to which that distribution is more or less equitable—to be a highly consequential and political process. In that sense, our perspective resonates with Gutiérrez's (2007) contention that power is a key dimension of equity. Indeed, referring to Gutiérrez's claim, Moschkovich (2011) reminds us that "the *power* dimension can involve examining voice in the classroom—for example, who gets to talk and how contributions are taken up (or not)" (p. 92). Research shows that certain groups tend to receive less cognitively demanding participation opportunities based on their gender (Sadker, Sadker, & Zittleman, 2009), race (McAfee, 2014), and immigration status (Planas & Gorgorió, 2004). Although such patterns may be unintentional, they are nonetheless problematic and immoral.

Overall, the literature suggests that discourse opportunities are central to learning but that not all students have access to such opportunities. In our view, researchers and practitioners need tools that can reliably, efficiently, and systematically illuminate these patterns of inequity. In order to attenuate inequity, though, we must first have the means to identify inequity. In the next section, we review existing observation tools that have sought to address this issue.

Treatments of Equity in Mathematics Classroom Observation Tools

The study of mathematics instruction has been formalized into a number of observation tools, including the Instructional Quality Assessment (IQA; Boston, 2012), the Mathematical Quality of Instruction (MQI) rubric (University of Michigan, 2006), and the Teaching for Robust Understanding of Math (TRU Math) rubric (Schoenfeld, 2014). Consistent with the preceding

---

[1] The focus on oral participation does not minimize the importance of other forms of participation (e.g., listening, gesture, writing, eye gaze), which also have been shown to be crucial to the learning process (Abrahamson & Sánchez-García, 2016; Schultz, 2009).

literature review, these tools attempt to capture the nature of classroom discourse. Moreover, we focus on these three tools because of their potential to speak to issues of equity, either directly (as in Goffney's (2010) MQE rubrics that extend the MQI), indirectly (as in the IQA's focus on participation in discourse), or through a hybrid approach (as in TRU's five dimensions—particularly the access dimension, which is focused on equity). Summaries of IQA (Boston, Bostic, Lesseig, & Sherman, 2015), TRU Math (Schoenfeld & the Teaching for Robust Understanding Project, 2016), and MQI (University of Michigan, 2006) can be found elsewhere; therefore, we focus on the equity components of the tools. In critically reviewing these tools, we recognize that all tools are designed to achieve particular goals and that no single tool could possibly capture everything that matters in a classroom. Thus, rather than discussing the quality of a tool in a vacuum, we find it more productive to consider the alignment between a tool and specific research aims.

We consider these tools from the perspective of three basic questions related to equity: (1) who participates, (2) what is the nature of that student participation, and (3) what opportunities do teachers make available to support that participation? These questions were informed by research linking participation, learning, and equity. Indeed, as we elaborate later, they undergird our own observation tool, EQUIP. Although all three tools appear reasonably effective at analyzing the second and third equity questions, we argue that they are not designed to capture information about the first equity question (who participates) at a sufficient level of specificity. In addition, because who participates and how they participate are treated separately, it is difficult to assess how meaningful opportunities to learn are actually distributed among students. In relation to the second equity question (the nature of student participation), all three tools emphasize discourse, focusing on how students explain and justify their mathematical ideas. They also describe the quality of mathematics during a lesson: TRU Math has dimensions that focus on "The Mathematics" and "Cognitive Demand" (see Appendix A), the IQA has rubrics for the academic rigor of a task, and the MQI includes a "Mathematical Features" rubric. In relation to the third equity question (teacher support for participation), the tools account for teacher moves intended to provide opportunities for students to engage with the mathematics. In the MQI, this is achieved through the addition of the MQE add-on module (see Appendix B) that emphasizes a teacher encouraging a "diverse array of mathematical competence" (Dimension G) and also expressing the expectation that "everyone will be able to do the work" (Dimension J).

TRU Math focuses on a teacher's use of formative assessment to elicit and build on student thinking (see Dimension 5 in Appendix A). The IQA focuses on the quality and explicitness of a teacher's expectations, teacher elicitation of ideas, and linking of student ideas.

Despite their capacity to capture important aspects of students' participation in mathematical discourse and teachers' efforts to support that participation, the tools have limitations with respect to the first equity question: who participates. The MQI (and the MQE variant) acknowledge the importance of student participation, but they do not include an explicit dimension on equitable participation. In the TRU conversation guide, it is noted that each dimension must take place for each student. In other words, TRU considers equity as a part of each dimension rather than as a separate an "add-on" dimension. However, without explicitly quantifying what counts as equitable participation, the interpretation of what is equitable is left to the observer.

The IQA contains a holistic rubric about student participation that establishes a quantitative benchmark for student participation and also records the actual number of students who participate in a lesson. The highest level of participation is described as "over 75% of the students participated consistently throughout the discussion" (see Appendix C). Although this is a strong step towards tracking participation, individual students or groups are not identified. Thus, even if the 75% benchmark were reached, it would be difficult to know if the 25% of students who were *not* participating came from historically marginalized groups (e.g., based on race or gender).

To summarize, EQUIP builds on existing work by addressing crucial aspects of equitable participation already valued by other tools. EQUIP links the second and third equity questions (the nature of student participation and teacher support), which helps answer how the quality of talk and opportunities to participate are distributed across individual students or groups (e.g., based on race or gender). Without the ability to disaggregate at the level of social identity groups and individual students, an observation tool is limited in how it can support efforts to attenuate inequity in classrooms. It is one thing to know that the breadth of student participation in a given classroom was narrow, but it is another to know that participation was limited to boys, White students, or language-dominant students. Information of the latter kind would hold great value to practitioners because it can cue them to notice inequities related to groups that have been historically marginalized in mathematics education. Our point here parallels a similar critique of the *Common Core State Standards for Mathematics*, which Bartell et al. (2017) argue are not "framed to support equity" because they lack "explicit attention to race, gender, class, and so forth" (p. 9).

We recognize that reducing socially constructed identities like race and gender to demographic categories can be problematic. Treating these complex sociopolitical constructs as variables flattens the fluid nature of identity and obscures the power-laden processes of positioning that organize social interaction (Davies, & Harré, 1990). Indeed, we ourselves have made similar arguments in prior work (Martin, Rousseau Anderson, & Shah, 2017; Shah & Leonardo, 2016). However, our rationale for doing so in this work is two-fold. First, we utilize Gutiérrez's (2002) notion of *strategic essentialism*, which she defines as "the process of deliberately categorizing people based on socially defined traits for the purpose of reaching higher (equity) goals" (p. 154). Strategic essentialism makes it possible to illuminate subtle patterns of inequity across larger data sets (e.g., girls receiving less cognitively demanding opportunities to participate across multiple lessons), which our methodology aims to do. For this reason, we conscientiously (but cautiously) utilize this approach and view it as a necessary step towards fostering equity for people from marginalized groups.

Second, we argue that the use of demographic categories in this particular context actually hews closer to how implicit bias functions. That is, implicit bias functions precisely by flattening complex social identities into static categories of race, gender, and other social markers (Staats, Capatosto, Wright, & Jackson, 2016). So, when using EQUIP to identify patterns of implicit bias, it actually makes more sense to use the teachers' perceptions of students' identities because it is the teachers' identifications that will drive their implicit bias. At the end of this article, we discuss how analytics generated through this approach can and should be considered in light of broader ideologies and discourses that might explain observed patterns of inequity. In the next section, we elaborate the theoretical perspective on equity that informs EQUIP and the equity analytics approach more broadly.

Equality as a Waypoint Towards Equity

Equity concerns matters of justice, or what is considered to be fair for individuals and groups (Green, 1983; Gutiérrez, 2008; Martin, 2003; Secada, 1989). Issues of equity in education can be understood in terms of the fair distribution of resources needed for learning (Esmonde, 2009). These resources are numerous and diverse: ranging from highly qualified teachers and access to advanced coursework (Darling-Hammond, 2010; Oakes, 2005) to curricula that value students' cultural practices and pedagogies that support the development of students' identities as capable learners (Bang, Warren, Rosebery, & Medin, 2012; Nasir, 2011). In this article, we focus on participation in classroom discourse as a key resource for learning.

From an equity perspective, all students deserve access to the particular types and levels of resources that they need to learn. However, for students from marginalized groups, there are often gaps between what a student needs and what a student gets. We define these gaps as inequities. But, how do we determine what a student needs, and who gets to decide? Given an education system with finite resources, what constitutes a fair distribution of resources? These questions pose a considerable challenge to observational approaches for studying equity and inequity in classrooms.

To move forward, we focus on equality as a waypoint towards equity. As Secada (1989) notes, "one of the most powerful constructs at the disposal of equity is equality" (p. 82). In the language of resources, equality is a condition in which all groups and individuals are apportioned the same type and level of resources. Technically speaking, then, any disproportionality in the allocation of resources would be an inequality. Importantly, the concepts of equality and inequality do not directly map onto the concepts of equity and inequity. Although these concepts are related, they are often erroneously conflated and treated as interchangeable (Espinoza, 2007; Secada, 1989).

For students from marginalized groups, it may be insufficient to receive the same resources as classmates from dominant groups. This is because all students do not enter a classroom with the same backgrounds. For instance, students' preparedness to learn new content is an artifact of past opportunities to learn, which are, in turn, mediated by institutional discrimination and longstanding structural inequities (Darling-Hammond, 1998; Milner, 2010). To account for this history of marginalization, ensuring fairness in opportunities to learn for students from marginalized groups might actually require allocating them more resources and different resources than students from dominant groups. Under certain circumstances, inequalities can actually be equitable (Samoff, 1996; Secada, 1989), or as Green (1983) puts it: "Inequity always implies injustice. Inequality does not" (p. 324). Broadly speaking, this means that conceptualizing equity in terms of fairness requires a historical orientation. Whereas equality focuses on the distribution of resources in the present moment, an equity perspective further involves accounting for the accumulated effects of histories of marginalization.

And yet, although equality is an insufficient principle for achieving equity, most would agree that, at minimum, students from marginalized groups deserve an equal share of resources for learning. That is, any shortfall relative to a proportional allotment would be unfair (i.e., an inequity). Although it may be difficult to determine the exact level and configuration of resources, it is possible to establish equality as a floor for equitable distribution. In the context of participation in classroom discourse, for example, this "floor" would be a proportional distribution of high-level questions based on demographic representation. Operationally, if Black students make up 21% of the students in the class, then at least 21% of high-level questions (i.e., an equal share) should go to Black students.

It is less clear that the same logic should apply for students from dominant groups who, by definition, have had greater historical access to opportunities to learn and therefore might not require a proportional level of resources to succeed. Thus, particularly for students from marginalized groups, equality can be understood as a waypoint towards equity. Equality is a necessary, albeit insufficient, baseline for equity. This idea provides a foundation for the equity analytics approach, which we elaborate in the next section.

Equity Analytics

We define equity analytics as a quantitative methodology for identifying misalignments in the distribution of resources. The core question driving equity analytics is: To what extent does the actual distribution of resources align or diverge from the distribution predicted based on demographic representation? Assuming that the resource of interest is quantifiable, comparing actual resource distributions to expected resource distributions is relatively straightforward. This approach has been commonly used to identify inequities in the high-tech workforce. For example, corporations like Google, Facebook, and Twitter have all publicized data on employee demographics, which have shown that women and certain racial groups remain underrepresented in engineering relative to their representation in the overall U.S. population (e.g., Google, n.d.).

Equity analytics can identify statistical equalities and inequities. Although we view this approach as promising, it does have limitations. Specifically, it cannot actually prescribe what resource distributions would be equitable. Even if there were consensus that people from marginalized groups deserve more resources as recompense for historical injustice, equity analytics cannot determine how much more they deserve. For example, equity analytics cannot tell us how many women, Black people, and dis/abled people should work as engineers at Google. We recognize that philosophical questions of this kind concern morality and values, and cannot be determined by quantitative analysis alone.

This issue takes on additional complexity when trying to make determinations about equity in classrooms. In part, the inability of equity analytics to make determinations about equity reflects the limitations of observational methods writ large. Classroom observation tools are third-party instruments that embody particular values and biases. The data that they generate reflect those values and biases, which may or may not align with the subjective perspectives of the students being observed. Consider a hypothetical in which analytics indicate that emergent multilingual students are receiving an equal share of adequate wait time, but when interviewed, those same students indicate that they feel that their teacher rushes them to respond to questions. How should this disconnect inform how we think about equity in classrooms?

By its nature, a classroom observation tool cannot gather data on students' subjective experiences of equity. Although the equity analytics generated by an observation tool might indicate equality in terms of students' opportunities to learn, the students themselves might not perceive the classroom to be an equitable space. No observation tool can or should supplant or supersede students' subjective experiences. Indeed, this is a facet of equity that observational methods cannot capture. We agree with Secada (1989) that, "If educational equity becomes equality of education, then arguments about justice are in danger of being recast as technical arguments about equality" (p. 74). The complexity of equity cannot be reduced to a simple mathematical calculation.

In our view, these are legitimate concerns. Although equity analytics alone cannot generate a comprehensive analysis of equity in classrooms, we argue that equity analytics can facilitate empirical research on equity, as well as support practitioners in making their teaching more equitable. In the next section, we describe EQUIP, a classroom observation tool that utilizes equity analytics.

EQUIP: Equity QUantified In Participation

EQUIP tracks quantitative patterns in student participation and the opportunities that teachers make available for students to participate in classroom discourse. As was mentioned earlier, EQUIP is predicated on three basic equity-related questions organized around student and teacher discourse: (1) who participates, (2) what is the nature of that student participation, and (3) what opportunities do teachers make available to support that participation? The first question concerns patterns in student talk: Who is participating in classroom discourse, and who is not? The second question acknowledges that in addition to who participates, equity is also a matter of how students get to participate (Wager, 2014). That is, opportunities to participate should be of a cognitively demanding nature (Stein, Grover, & Henningsen, 1996). A classroom in which all students participate but only in low-level, Initiate-Response-Evaluate (IRE) sequences (Cazden, 2001; Mehan, 1979) would be considered less equitable because students are not afforded opportunities to participate as active sense makers. The third question concerns what teachers do to support student participation. This question recognizes that teachers play a critical role in orchestrating classroom discourse and in distributing opportunities to participate in discourse (Ing et al., 2015; Parks, 2010; Turner et al., 2013).

EQUIP cross-tabulates certain dimensions of classroom discourse with demographic information to identify equalities and inequities in how the amount and nature of student participation are distributed. For example, does the teacher only ask high-level questions to the White and Asian students, boys, or English-dominant students? In equitable classrooms, the quality of students' opportunities to participate would not skew dramatically towards any groups or individuals, especially those belonging to historically dominant groups.

Versions of this approach have been used in prior studies on gender marginalization in classrooms (Sadker et al., 2009) and in practitioner inquiry (Neumann, 2014). However, equity analytics has not been systematized in a valid and reliable classroom observation tool that can generate a synthetic analysis of multiple facets of classroom discourse. Of the various facets of classroom discourse, EQUIP collects data on seven dimensions: Three dimensions focus on the amount and nature of student talk, and four dimensions focus on teacher moves consequential for students' opportunities to participate. We elaborate these dimensions in the section Validity of EQUIP.

Whereas existing classroom observation tools generally consist of descriptive rubrics, EQUIP relies on the tabulation of frequently occurring discursive events. Although rubrics can provide holistic portraits of classroom activity, they require considerable interpretation, which can make it difficult to achieve reliability. As we describe, the seven dimensions of EQUIP were chosen, in part, because they are both relatively straightforward to observe and require minimal inference to code. This was a purposeful design decision in the development of EQUIP intended to facilitate both the training of new coders and widespread use of the tool. Further, this allows an observer using EQUIP to take a less judgmental role (i.e., tabulating events rather than locating performance on a rubric), which we view as key to working in partnership with teachers to address issues of equity. This will support our future work using EQUIP as a real-time observation tool that is capable of providing immediate feedback to teachers for use in professional development.

To reiterate, tabulating events also makes it possible to disaggregate the frequency of those events by demographic group and by individual student (i.e., investigate the "who participates" question). This disaggregation is critical to understanding how opportunities to learn are distributed across socially constructed markers of difference (e.g., gender, race). Further, data on individual students are useful because group-level analysis alone can obscure within-group variation (Secada, 1989). For example, although data might show a proportional amount of participation from Black boys, this does not necessarily mean that all Black boys in the class are participating in equal amounts.

A final key design feature of EQUIP is modularity. The version of EQUIP presented in this article is one possible instantiation of equity analytics. The seven dimensions of classroom discourse included in this version of EQUIP are not intended to be exhaustive—as we explain, we identified these dimensions based on decades of research showing their relevance to equity issues in classrooms. Other researchers and practitioners, however, might identify additional dimensions of classroom discourse relevant to equity. The modular nature of EQUIP allows for these dimensions to be added to the current version of EQUIP or for existing dimensions to be removed.

The flexible nature of EQUIP makes the tool potentially adaptable to a variety of classroom types. This also applies to variations in demographics. For example, consider the case of a racially homogeneous classroom (e.g., all Black students). This is an important case in light of research showing trends toward resegregation between and within U.S. schools (Kalogrides & Loeb, 2013; Orfield & Lee, 2007). Although analyzing patterns based on traditional racial categories would not make sense in that context, an EQUIP user might instead analyze patterns based on new categories related to skin tone (e.g., darker skinned Black students vs. lighter skinned students). Indeed, research has documented evidence of skin color hierarchies in which lighter skin tones are unfairly privileged over darker skin tones among members of the same racial group (Hunter, 2007). A similar approach can be used in the case of gender; an EQUIP user can create and track on any number of gender categories, thereby acknowledging the limitations of the commonly accepted binary model of gender. In that sense, researchers and practitioners can customize EQUIP based on their needs and local contexts while attending to the fluid quality of social identities.

Using EQUIP

In this article, we present EQUIP as a research tool for analyzing video records of classroom discourse. The primary unit of analysis in EQUIP is the participation sequence. A participation sequence is a string of talk by a given student—often with a teacher—that is uninterrupted by another student (i.e., a new participation sequence begins each time a new student participates). Thus, participation sequences may consist of multiple turns by the same student with back-and-forth between the student and the teacher. For the sake of brevity, we henceforth refer to participation sequences simply as sequences.

A coder begins by identifying sequences. Each sequence is associated with a particular student and then coded along seven dimensions (described in the next section). In addition, a coder creates a spreadsheet with demographics (e.g., gender, race) for the students in the class. To generate equity analytics, the coded transcript and demographic data are combined using the statistical package R. This allows a user to generate analytics related to any aspect of coding and any combination of demographic variables.

Validity of EQUIP

We establish the validity of our tool by describing the seven dimensions of EQUIP, the levels at which they are coded, and their connections to the literature (see Table 1 at the end of this section for a summary). As described earlier, EQUIP's focus is consistent with the research literature and existing classroom observation tools. This consistency helped establish the *construct validity* of EQUIP (Peter, 1981).

We also iteratively incorporated feedback from nine one-on-one consultations with expert colleagues, three conference presentations, and a series of equity-focused workshops with practicing teachers. Moreover, we conducted a thorough review of the literature supporting each of EQUIP's dimensions. This helped establish the *content validity* of EQUIP (Haynes, Richard, & Kubany, 1995). We now elaborate the research base for each dimension.

Discourse Type

EQUIP distinguishes two types of classroom discourse: content and logistical. *Content-oriented discourse* involves disciplinary practices consequential for student learning, such as explanation (Driver, Newton, & Osborne, 2000). *Logistical discourse* concerns other talk, such as talk regarding classroom norms (Michaels et al., 2010). Both types of discourse are consequential for equity, albeit in different ways: content-oriented discourse determines who gets to "do the work" in class (leading to learning), whereas logistical discourse about classroom rules, for example, helps structure opportunities for all students to engage with disciplinary content. Logistical discourse matters because when classroom norms are left implicit, students may be denied opportunities to participate in content-oriented discourse.

Length of Student Talk

Powerful classrooms are ones in which students generate the ideas that drive classroom discourse (Engle & Conant, 2002). For instance, the *Common Core State Standards for Mathematics* emphasize that students need to construct arguments, critique others' reasoning, and attend to precision (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA & CCSSO], 2010). This requires that students have opportunities for elaborated talk (e.g., Soter et al., 2008). Opportunities for elaborated contributions are often limited by IRE structures, which generally elicit responses of only 1–4 words, (Cazden, 2001; Mehan, 1979). The categories of 5–20 words and 21+ words roughly correspond to a single sentence and multiple sentences. The latter category can be understood as an indicator of higher level discourse because in highly teacher-centered, IRE-based classrooms, students may rarely get to utter a second sentence during a given turn. Further, opportunities to talk at length are particularly important for supporting the language development of emergent multilingual students (M. Boyd & Rubin, 2002).

Type of Student Talk

When students explain their ideas, it helps them engage with tasks at a higher level of cognitive demand (Henningsen & Stein, 1997), which ultimately supports deeper learning (e.g., Chi, De Leeuw, Chiu, & Lavancher, 1994; Lombrozo, 2006). To characterize the quality of student talk, we draw from Braaten and Windschitl's (2011) approach for classifying verbal statements, which resonates with Henningsen and Stein's (1997) framework for cognitive demand. Specifically, this approach classifies statements on three levels: *why* statements, *how* statements, and *what* statements (Braaten & Windschitl, 2011).

Teacher Solicitation Method

Teachers solicit participation from students in a variety of ways. Some methods, such as randomized calling (Tanner, 2013), can help equalize participation and also signal to students that contributions from all members of the class are valued. Without using explicit methods to promote equity, teachers can inadvertently favor some students over others because bias is often implicit (Sadker et al., 2009; Staats et al., 2016). Finally, students may participate without being called on at all by the teacher. On the one hand, this may reflect a classroom in which students have agency to take ownership over their own learning (Engle, 2012). On the other hand, the absence of a formal system for soliciting participation can result in inequitable distributions of participation opportunities.

Wait Time

To broaden participation, a teacher may explicitly slow down to give students more time to think before responding to a question. Rowe's (1986) seminal research shows that the critical threshold for this "wait time" is approximately 3 seconds. Beyond this threshold, research shows a number of benefits for learners, including an increase in the length and quality of student talk, an increase in student questions, and an increase in student–student talk (Rowe, 1986). Wait time signals to students that deep thought is more important than a quick response. This is particularly important in disciplines like mathematics because the perception that speed is all that matters can have a detrimental impact on students (Schoenfeld, 1988). A lack of wait time can be a serious equity issue because it can privilege the participation of the fastest students in a classroom.

Quality of Teacher Solicitation

Teachers can modify the cognitive demand of a task through their use of questioning (Henningsen & Stein, 1997). Lower level questions often remove the challenge from a task by reducing its cognitive demand. Alternatively, teachers can ask questions that raise the cognitive demand of a task by pressing students to think more deeply (M. Boyd & Rubin, 2002). To be sure, we recognize that the deployment of higher level questions will not automatically lead to greater equity. Indeed, research shows that pedagogical moves may not be taken up in the same way or be equally productive for all students in a classroom (Lubienski, 2000; Parks, 2010). In general, though, we consider it valuable to track patterns in the quality of teacher solicitations.

Teacher Evaluation of Student Ideas

To engage productively with a discipline, students need opportunities to take up positions of authority (Engle, 2012). When students lack access to such opportunities, they may develop unproductive beliefs about themselves and the discipline (Schoenfeld, 1988). One way that the locus of authority is communicated to students is through teacher evaluations of student ideas. For example, in many classrooms the teacher acts as the primary arbiter of right and wrong answers. In such classrooms, the message to students is that authority lies with the teacher as opposed to with the students or the discipline. Our coding scheme involves indicating *yes* when a teacher evaluates an idea and *no* when a teacher leaves the correctness of the idea open, allowing for other students to evaluate it (or not).

Table 1

Dimensions of EQUIP

Dimension	Levels	Core literature

Discourse type	Content Logistics	Driver et al., 2000; Michaels et al., 2010; Yackel & Cobb, 1996
Length of student talk	21 or more words 5–20 words 1–4 words	M. Boyd & Rubin, 2002; Cazden, 2001; Engle & Conant, 2002; Hufferd-Ackles et al., 2004; Mehan, 1979; Michaels et al., 2010; NGA & CCSSO, 2010
Type of student talk	Why How What Other	Braaten & Windschitl, 2011; Chi et al., 1994; Henningsen & Stein, 1997; Stein, Engle, Smith, & Hughes, 2008; Lombrozo, 2006
Teacher solicitation method	Random selection Called on NOT called on	Engle, 2012; Sadker et al., 2009; Tanner, 2013;
Wait time	More than 3 seconds Less than 3 seconds N/A	Rowe, 1986; Schoenfeld, 1988
Type of teacher solicitation	Why How What Other N/A	M. Boyd & Rubin, 2002; Braaten & Windschitl, 2011; Chi et al., 1994; Henningsen & Stein, 1997
Teacher evaluation of student ideas	Yes No	Engle, 2012; Schoenfeld, 1988

Reliability of EQUIP

We now discuss our procedures for establishing EQUIP as a reliable observation tool.

Data Sources and Research Context

Video data were collected during summer 2014 in a 2-week elementary mathematics program for students entering 5th grade in the fall. The teacher, Ms. J, had decades of teaching experience and also held deep commitments to equity. Ms. J identified as a White woman. This site was further appealing because the program purposefully recruited students struggling in mathematics, often from historically marginalized groups (see Table 2).

Table 2

Student Demographics by Race and Gender (N = 30)

	Girls	Boys	Total
Black	10	11	21 (70%)
Latinx	3	2	5 (17%)
White	2	2	4 (13%)
Total	15 (50%)	15 (50%)	30 (100%)

Note. The racial and gender identifications were determined by the teacher, Ms. J, based on her extensive interactions with students and their families during the course. However, we use the *x* sign in the word *Latinx* to acknowledge that all people do not identify with the traditional gender binary.

The program consisted of 30 hours of instruction (3 hours per day over 10 days). Content focused on fraction representations and norms for engaging in mathematical discourse. A typical class session involved whole-class discussion (WCD), small-group work, and individual work. The forthcoming analysis focuses on WCDs in Ms. J's class because this was where teacher–student interaction was most readily observable. The majority of WCDs (70 discussions, 8.5 hours) were content oriented (i.e., about mathematics), whereas the other WCDs (23 discussions, 1.25 hours) were logistical (i.e., about classrooms norms and academic expectations). Transcriptions of all WCDs comprise the data corpus that was analyzed.

Establishing Reliability

Reliability was established using a team of five raters working from a common codebook created from prior work. The team consisted of the two coauthors (mathematics education researchers) and three student research assistants (two graduate students and one undergraduate student). Although demographics were not formally gathered, the team was diverse in terms of race and gender. Thus, the team approached coding from a variety of perspectives, but all team members were focused on equity and mathematics education.

To begin, the team created content logs of all 10 days of instruction (over 1,300 sequences). To establish reliability, we sampled 120 minutes of WCD, which was slightly more than 20% of the 9 hours and 43 minutes in the entire data corpus. We created a random sample that had the same proportion of content-oriented and logistical discourse as the whole data corpus; this resulted in approximately 105 minutes of content-oriented discourse and 15 minutes of logistical discourse. Our benchmark was to achieve 80% agreement by all five raters on each dimension over a sufficient corpus of data (Grossman et al., 2010). To reach this benchmark, we added an additional 35 minutes of video to the sample coded for reliability because we had not reached 80% on the original sample. Overall, the sample spanned 7 of the 10 days of instruction, thereby constituting a representative sample.

Coding for reliability proceeded in five phases. Each phase involved coding approximately 30 minutes of video with its corresponding transcript. After coding each phase, reliability was calculated and the team discussed areas of disagreement, supplementing the codebook with additional clarification as needed. Disagreements were only discussed in the general sense; no effort was made to resolve disagreements on a line-by-line basis.

Agreement was computed as a binary *yes* or *no*. That is, an agreement was recorded only if all five raters agreed on the assignment of a code. We first computed agreement on the identification of sequences. When calculating the percentage of agreement, we included all lines of transcript that were identified as new sequences by at least one rater. Agreement on other codes was calculated in two ways. First, we calculated agreement for all coders who coded a given sequence, which could be less than five if not all coders identified that particular sequence. These results are provided in Table 3. We also calculated agreement for all of the sequences in which there was total agreement on the identification of the sequence, meaning that all five raters coded them. These results, which are not included here, were consistent with the results in Table 3; both approaches resulted in greater than 80% agreement on all codes. In calculating agreement on the dimensions of EQUIP, each code was considered individually.

In addition to providing measures of pure agreement, we computed interrater reliability using Krippendorff's alpha (Hayes & Krippendorff, 2007). Krippendorff's alpha generalizes a number of other interrater reliability statistics (including Cohen's kappa) to be appropriate for more than two raters and for multiple code levels.

Results of Reliability Procedure

After five phases of coding, we reached 80% agreement on all dimensions over the sampled corpus of data. As Table 3 shows, we achieved greater than 80% agreement for all dimensions in each phase with only four exceptions: Wait Time during Phase IV (75%) and Type of Teacher Solicitation during Phases II, III, and V (70–80%).

Table 3

Percentages of Agreement per Dimension per Phase

Dimension	Phase I	Phase II	Phase III	Phase IV	Phase V	Overall
Sequences	82.6%	85.7%	92.8%	94.9%	94.7%	89.5%
Discourse type	98.8%	98.9%	98.6%	96.4%	100.0%	98.7%
Solicitation method	96.5%	95.4%	94.2%	92.9%	96.0%	95.2%
Wait time	84.7%	87.4%	84.1%	75.0%	80.0%	82.8%
Length of student talk	95.3%	92.0%	91.3%	94.6%	93.3%	93.3%
Type of student talk	88.2%	86.2%	81.2%	96.4%	80.0%	86.0%
Type of teacher solicitation	89.4%	72.4%	72.5%	94.6%	76.0%	80.4%
Explicit evaluation	90.6%	87.4%	88.4%	91.1%	86.7%	88.7%

Results for Krippendorff's alpha are given in Table 4. All results are over 0.8, which is considered good reliability, the highest category that can be achieved (Carletta, 1996). These results confirm the results for pure agreement.

Table 4

Krippendorff's Alpha for Seven Dimensions (N = 359 Cases, 5 Raters)

Dimension	Discourse type	Solicitation method	Wait time	Length of student talk	Type of student talk	Type of teacher solicitation	Explicit evaluation
Alpha	0.976	0.934	0.844	0.942	0.989	0.875	0.835

Despite excellent agreement, there were some minor discrepancies in coding. Still, we found no evidence that a single rater was systematically interpreting a number of codes differently from the other raters. We now provide additional information about these four areas of discrepancy. For Wait Time, Phases IV and V had the lowest levels of agreement: 29 discrepancies in total across both phases. In 12 of the 29 (41%), there was a single rater that disagreed with the other four. Among these 12 instances, 7 involved the same rater disagreeing with the other four raters. This means that 24% of the disagreement can be accounted for by differences in how this particular rater interpreted the Wait Time code compared to other raters' interpretations, which tended to align.

For Type of Teacher Solicitation, there were 24 disagreements during Phase II. Of these 24 disagreements, there were 16 instances (66.6%) in which only a single rater disagreed with the other four. Across Phases I and II, 17 of 33 (52%) disagreements on Type of Teacher Solicitation involved difficulties in distinguishing between *N/A*, *other*, and *what*. This issue most commonly arose during moments in the transcript when a student spoke without their participation being solicited by a teacher. In these cases, some raters coded Type of Teacher Solicitation as *N/A* (indicating that there was no Teacher Solicitation), whereas the other raters coded Teacher Solicitation as *other*.

During Phase III, there were 19 disagreements for Teacher Solicitation. The majority of these disagreements (10 of 19 times, or 53%) are attributable to a single rater disagreeing with the other four raters. There were also six instances in which two raters coded Type of Teacher Solicitation as *other*, and the other three raters coded Teacher Solicitation as *why*. This was due to ambiguity surrounding whether the teacher was explicitly asking for an explanation or not.

During Phase V, seven disagreements about Type of Teacher Solicitation came during a single segment of transcript (Lines 111–134) for which two raters repeatedly applied a *what* code, and the other three raters applied a *why* code. This accounts for 42% of the disagreements. Other than this repeated disagreement, there appeared to be no systematic discrepancies.

Overall, most disagreements occurred with only a single rater (not necessarily the same rater) disagreeing with the other four raters. Thus, despite a small number of differences, we achieved a high level of agreement on a large corpus of data (over 1,300 sequences), ultimately achieving greater than 80% agreement on all dimensions. Before presenting the equity analytics generated by applying EQUIP to the full data corpus, we analyze a vignette from Ms. J's classroom. After qualitative analysis of the vignette, we show how equity analytics generated by EQUIP can augment our understanding of equity and inequity in Ms. J's classroom.

A Look Inside Ms. J's Classroom

In general, students in Ms. J's classroom were presented with rich mathematical tasks and engaged in discursive practices central to doing mathematics. Ms. J established strong sociomathematical norms in her classroom and continually reinforced these norms through classroom discussions and assignments. The following vignette took place midway through the program on Day 5, at which point classroom norms and expectations had largely been established. The lesson that day concerned the number line. Ms. J's goal was to orient students to aspects of the number line that they previously might not have problematized. The task that students were given involved a number line on which the numbers 0, 1, and 2 had been marked. Ms. J then added an additional unlabeled mark in the position where the number 4 would be. In previous whole-class discussions, the class had discussed the idea that equal distances on the number line represent equal differences between numbers. With this task, students were asked to determine which number should be at the unlabeled mark and also to predict which number a younger, second-grade student might think would go there and why. The following transcript involves two students, Bernardo (a Latinx boy) and Monique (a Black girl):

Line 1	Bernardo:	Four.
Line 2	Ms. J:	Okay, and why do you think it's four, Bernardo?
Line 3	Bernardo:	Because the distance from two and four—
Line 4	Ms. J:	—Can you put your hand down? [to another student]
Line 5	Bernardo:	There should be a number in the middle of two and four.
Line 6	Ms. J:	Okay so you think there should be a number here? And what number do you think that should be?
Line 7	Bernardo:	Three.
Line 8	Ms. J:	And why do you know three goes there, Bernardo?
Line 9	Bernardo:	Because all the numbers have the same distance.
Line 10	Ms. J:	Excellent. Now let's take off the answer that Bernardo gave for a minute, and who can tell me what number you think a second grader might put there if a teacher showed that, or if they saw a number line? Monique, what do you think somebody might put?
Line 11	Monique:	Three.
Line 12	Ms. J:	And why do you think somebody might put three there?
Line 13	Monique:	Because one—because after two is three.
Line 14	Ms. J:	Right, so they wouldn't be thinking about distance, they would just go in order. Okay, so Bernardo is right, this is four.

After Bernardo provides the correct answer (Line 1), Ms. J asks Bernardo to explain and justify his thinking (Line 2). A similar pattern occurs on two other occasions later in the vignette

(Lines 7–8 and Lines 11–12). The message to students is that correct answers alone are insufficient. When a classmate raises their hand to speak while Bernardo is speaking, Ms. J asks them to put down their hand (Line 4). One interpretation of this move is that Ms. J wants the rest of the class to focus on Bernardo's thinking. Overall, Bernardo and Monique are taking the lead in the classroom discussion, and Ms. J supports them by probing their sense making about the mathematics. Both Bernardo and Monique are members of racial and gender groups historically marginalized in mathematics classrooms, so their participation in mathematical discourse is noteworthy.

Although this vignette suggests that Ms. J's classroom may have been an equitable space for mathematical learning, the vignette also raises questions about the types of claims that can be made about equity and inequity across all 10 days of instruction. How typical was the kind of classroom discourse captured in this vignette? To what extent did other Black and Latinx students or girls also have opportunities to participate in the ways made available to Bernardo and Monique? And what about Bernardo and Monique themselves: Was this vignette typical of their experiences in Ms. J's class, or was this an outlier amidst a broader pattern of marginalization? Next, we demonstrate how equity analytics generated by EQUIP can complement and add context to such qualitative analysis as well as illuminate subtle nuances in equity and inequity that are difficult to detect without considering patterns across the entire data corpus.

Quantitative Results

To generate equity analytics for Ms. J's classroom, the entire corpus of WCDs was coded by the three student researcher assistants from our coding team. The quantitative results presented here are organized into four parts: (1) classroom-level results, (2) group-level results disaggregated by gender and race, (3) results disaggregated at the level of individual students, and (4) limitations of the quantitative results.

Classroom-Level Results

A total of 1,340 sequences were coded for the entire 10 days of classroom activity. Figure 1 provides a histogram of student participation. The mean number of sequences per student was 44.7 (*SD* = 26.5), approximately 4–5 sequences per student during each day of instruction. The tails of the histogram show that there were several students who rarely participated (0–20 sequences) and two students who participated far more frequently (101–140 sequences) than the rest of the class.

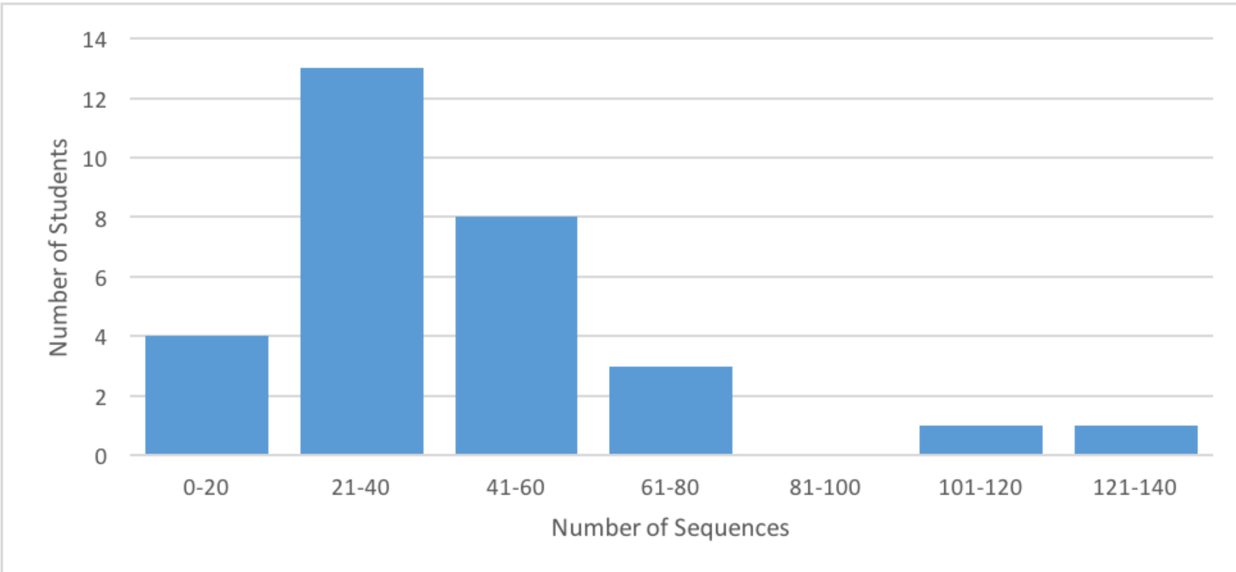

Figure 1. Histogram of student participation.

Figure 2 shows the distribution of sequences broken out by individual student, which provides additional perspective on the vignette of Ms. J's classroom. The focal student in the vignette, Bernardo, only participated in 18 sequences over the entire 10 days of instruction, the least amount of participation amongst all boys in the class. This suggests that his participation in the earlier vignette was not representative of his participation during most class sessions. In contrast, the other student in the vignette, Monique, had one of the highest levels of participation in the class (the fourth highest amongst girls).

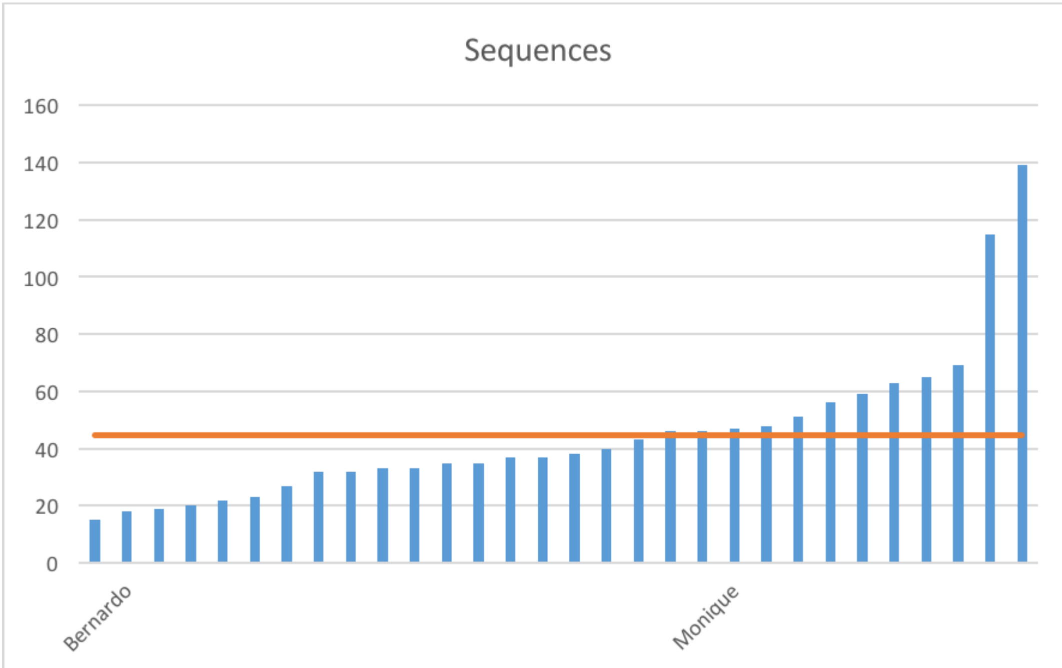

Figure 2. Number of sequences by student (mean = 44.7).

Table 5 summarizes student participation. Nearly 80% of sequences occurred during content-oriented WCDs, which comprised 87% of the instructional time in WCDs. Table 5 also shows that 24% of student participation involved mathematical explanation (i.e., *why* type of student talk). The *other* category accounts for sequences in which students asked questions, classroom norms were discussed, or there was other talk that did not directly make a mathematical contribution. We see also that a total of 67% of student participation was longer than short, 1–4 word contributions. These analytics indicate a classroom that is organized around participation in meaningful discourse.

Table 5

Classroom and Student Talk Characteristics

Discourse type	Type of student talk	Length of student talk
Content (79.5%) Logistics (20.5%)	Why (24%) How (3.0%) What (29.3%) Other (43.7%)	21+ words (16.8%) 5–20 words (49.8%) 1–4 words (33.4%)

Table 6 summarizes teacher talk. Ms. J solicited the majority of student participation (76.2%) by calling on students; no randomized calling methods were observed. When Ms. J called on students, more than half of the solicitations (53.6%) had a wait time greater than 3 seconds, indicating that she recognized students' need for time to think about the content being discussed. Consistent with results for student participation, nearly one fourth of solicitations (22.2%) focused on prompting the students to explain or justify their thinking (i.e., *why* types of solicitations). Compared with many mathematics classrooms, which tend to be organized around the IRE discourse pattern typically driven by *what* solicitations, we consider the proportion of *why* solicitations in Ms. J's classroom to be relatively high. Finally, when students did participate, the majority of the time (80%), the teacher did not explicitly evaluate students' mathematical contributions. In other words, Ms. J allowed authority to reside with the students or the mathematics rather than centering herself as the arbiter of correct and incorrect ideas.

Table 6

Teacher Talk Characteristics

Solicitation method	Wait time	Type of solicitation	Evaluation
Called on (76.2%) Not called on (23.8%)	3+ seconds (53.6%) 0–3 seconds (22%) N/A (24.4%)	Why (22.2%) How (3.2%) What (28.0%) Other (29%) N/A (17.6%)	No (80.0%) Yes (20.0%)

Overall, the equity analytics related to classroom-level indicators align with the qualitative analysis of the vignette presented earlier. Ms. J's classroom appears to be a place in which students were afforded opportunities to participate in mathematics in meaningful ways and often took advantage of those opportunities. And yet, this does not mean that all students—particularly those from historically marginalized groups—participated in the same ways or in the same frequency. As we noted above, Bernardo rarely participated in mathematical discussions, so his participation in the earlier vignette was an outlier. In the next section, we analyze patterns of participation and participation opportunities disaggregated by gender and race. To compare actual to expected participation (based on classroom demographics), we introduce a metric called the *equity ratio*.

Group-Level Results and the Equity Ratio

The equity ratio is defined as the ratio of *actual participation* to *expected participation* for a group of students along a particular dimension of classroom discourse. Actual participation is determined by classroom observation using EQUIP. Expected participation is what one would predict based on a group's demographic representation in a classroom.

The equity ratio can fall into three categories: greater than one, equal to one, or less than one. A ratio greater than one means that the actual participation of a group exceeds what one would predict in an "equal" classroom (i.e., the group's participation is overrepresented along that dimension). A ratio equal to one means that actual and expected participation are the same. Finally, a ratio less than one means that a group's participation is less than what one would predict in an "equal" classroom or that the group's participation is underrepresented along that dimension.

Here we consider how three dimensions from EQUIP (Type of Student Talk, Length of Student Talk, and Type of Teacher Solicitation) were distributed along the demographic variables of gender, race, and the intersection of gender and race. EQUIP generates a large number of quantitative metrics, and it is crucial that a user does not simply apply EQUIP to all possible dimensions of classroom discourse and demographic variables to look for significant results (i.e., "*p*-hacking"). To avoid this, we recommend that researchers and practitioners generate a priori hypotheses and focus on an appropriate subset of EQUIP codes.

Figure 3 indicates that the equity ratio was close to 1 for all types of student talk; any differences were not significant, $\chi^2(3, N = 1340) = 5.84$, $p = 0.12$. Figure 3 also describes the length of student talk by gender. Once again, the equity ratio was close to 1. However, boys engaged in lengthier discourse than girls (as indicated by more "21+ word" sequences), and the differences were significant, $\chi^2(2, N = 1340) = 18.1$, $p = 0.00012$, Cramer's V = 0.12, indicating a small effect size. Following up on the overall significance of the chi-square test, we conducted post hoc analyses of the standardized adjusted residuals associated with "21+ words." There were values of 4.54 (boys) and -4.38 (girls), which cross the threshold of +/- 2, indicating that the individual differences were statistically significant (Sharpe, 2015).

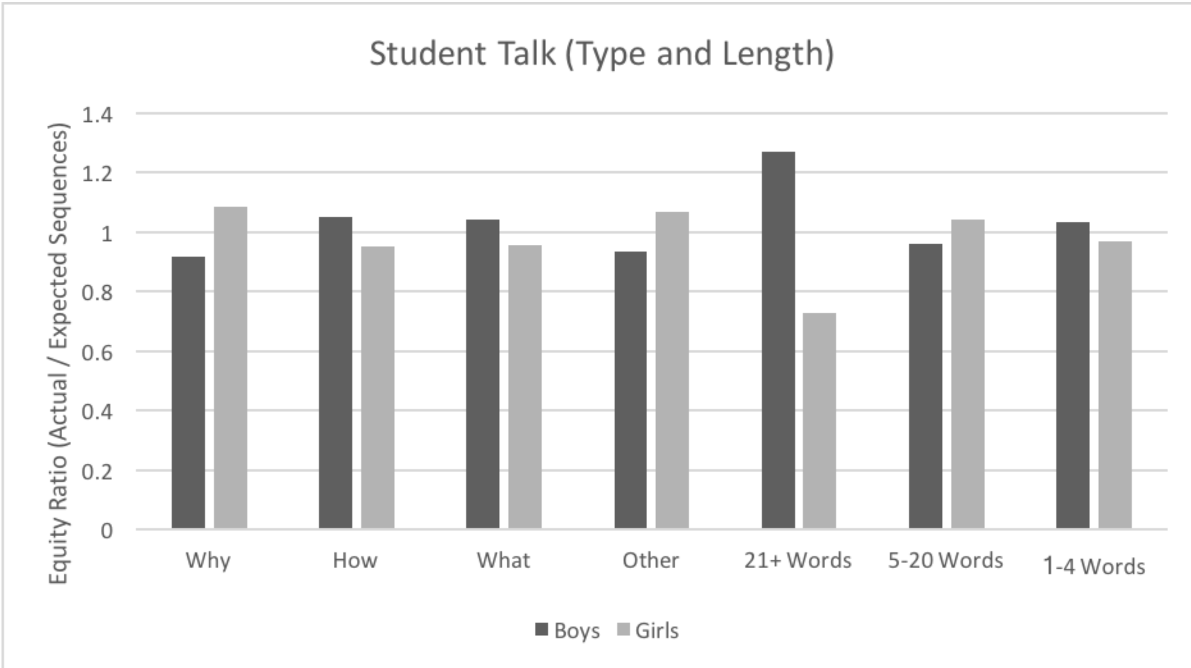

Figure 3. Student talk (type and length) by gender.

Figure 4 describes teacher solicitation type by gender. Once again most of the values were near 1, and the results were not significant, $\chi^2(4, N = 1340) = 7.00$, $p = 0.13$. Overall, these results suggest that there were no glaring inequities in classroom discourse by gender along these dimensions.

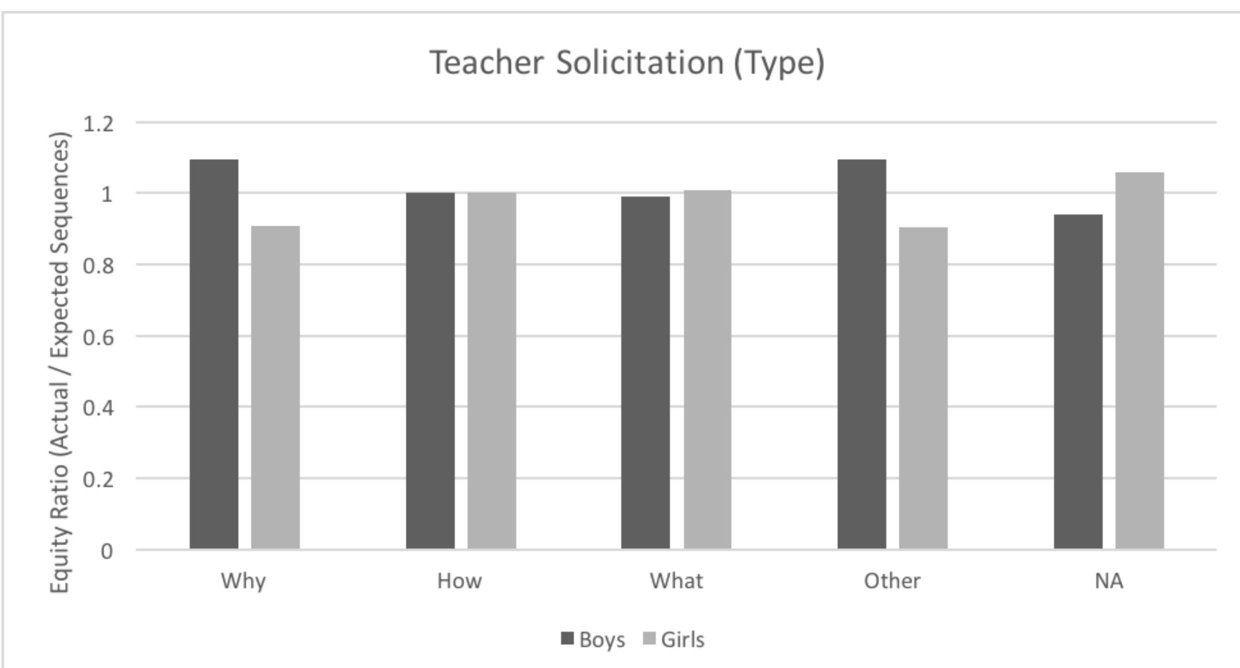

Figure 4. Teacher solicitations (type) by gender.

Next, we consider the analytics by race. Figure 5 shows the breakdown of student talk by race. The equity ratio for Black students across all dimensions was close to 1, which suggests proportional levels of participation. However, the analytics also reveal subtle inequities. Although all students had opportunities to engage in mathematical explanations, White students provided significantly more *why* statements than one would predict based on demographics. We also see that across categories, Latinx students participated less overall than one would predict. These differences were significant, $\chi^2(6, N = 1340) = 43.9$, $p = 7.6 \cdot 10^{-8}$, Cramer's V = 0.13, indicating a small effect size. Standardized adjusted residuals of -2.48 (Latinx) and 3.07 (White) indicate that these results were significant for *why* statements (Sharpe, 2015). Regarding length of talk, White students provided the majority of the extended contributions, whereas Latinx students were less likely to do so. These differences again were significant, $\chi^2(4, N = 1340) = 57.1$, $p = 1.2 \cdot 10^{-11}$, Cramer's V = 0.146, indicating a small effect size. Standardized adjusted residuals of -3.51 (Latinx) and 4.50 (White) indicate that these results were significant for "21+ words" (Sharpe, 2015).

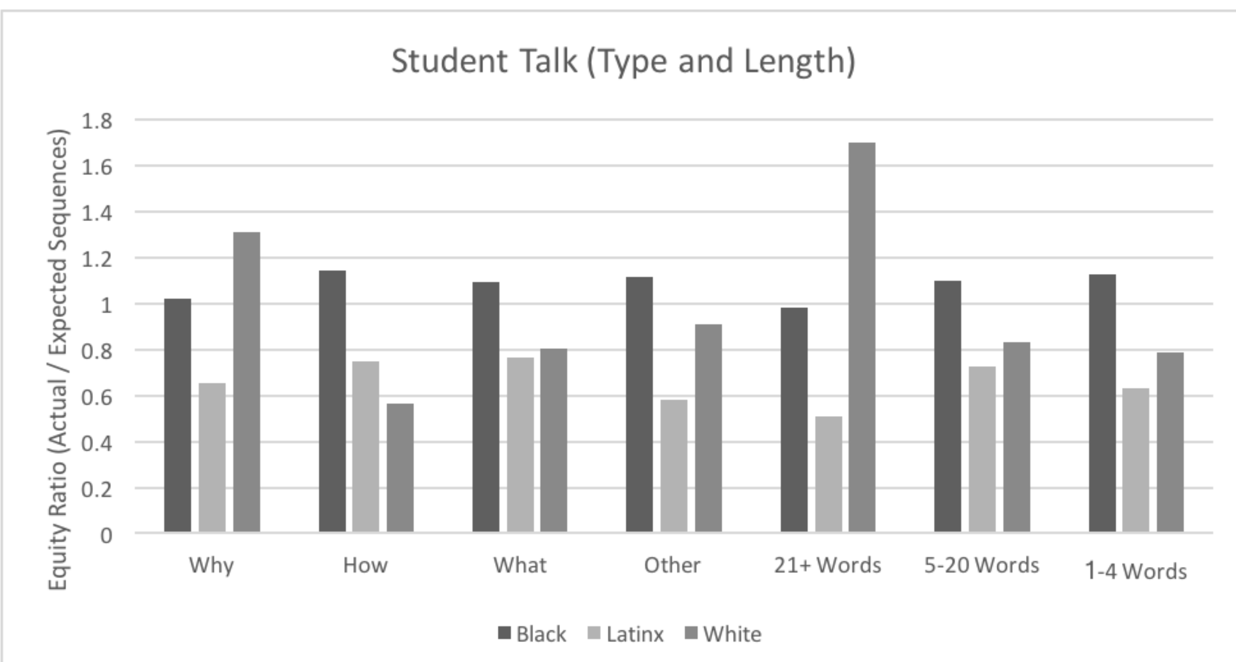

Figure 5. Student talk (type and length) by race.

Figure 6 shows teacher solicitation type by race. The results here were relatively consistent with those shown in Figure 5: Black students received proportional opportunities to participate, White students received a disproportionately greater number of *why* questions, and Latinx students were underrepresented in every category. Figure 6 adds a new level of depth with the *N/A* category, which occurred when students spontaneously offered ideas to the discussion; Black students were most likely to participate in this way. The differences shown in Figure 6 were significant, $\chi^2(8, N = 1340) = 52.58$, $p = 1.4 \cdot 10^{-8}$, Cramer's V = 0.140, indicating a small effect size. Standardized residuals of 5.94 (for Black students) indicated that these results were individually significant (Sharpe, 2015).

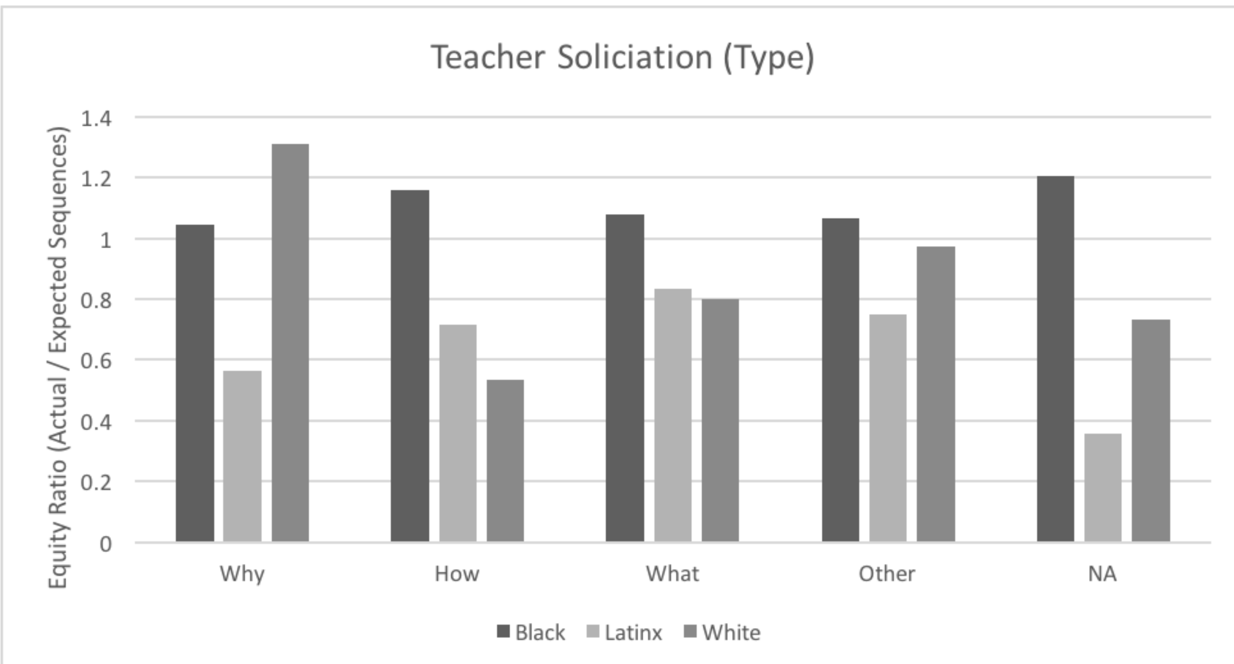

Figure 6. Teacher solicitation (type) by race.

Finally, we consider the intersection of race and gender. Figure 7 offers additional nuance to the race-only analytics shown in Figure 5. The race-only analytics showed that participation of Black students overall was found to be proportional, and here the intersectional analytics show that there were also negligible differences in participation between Black boys and Black girls. The one dimension in which there was a significant difference was the length of student talk: Black girls were less likely to make extended contributions than Black boys, $\chi^2(10, N = 1340) = 106.0$, $p = 3.3 \cdot 10^{-18}$, Cramer's V = 0.20, indicating a small to medium effect size. This was confirmed by looking at standardized adjusted residuals of 4.49 (Black girl) and 4.23 (Black boy; Sharpe, 2015). Next, the race-only analytics showed that White student participation was disproportionately high with *why* level contributions and lengthier talk. Here the intersectional analytics show that although participation from White girls was proportional for these dimensions, it was White boys that accounted for the overrepresentation of White student participation. Similarly, the intersectional analytics show that the disproportionately lower participation by Latinx students overall is best explained by Latinx boys, who consistently participated at lower rates. Fisher's exact test was used to account for small expected values in the contingency table. The results were significant ($p = 8.26 \cdot 10^{-6}$), Cramer's V = 0.13, indicating a small effect size. Post hoc analyses showed standardized residuals of magnitude 3 or greater (accounting for the large contingency table size) for *what*, *why*, and *other* and for 21+ and 5–20 words, indicating that they were all significantly different from what one would expect by chance (Sharpe, 2015)

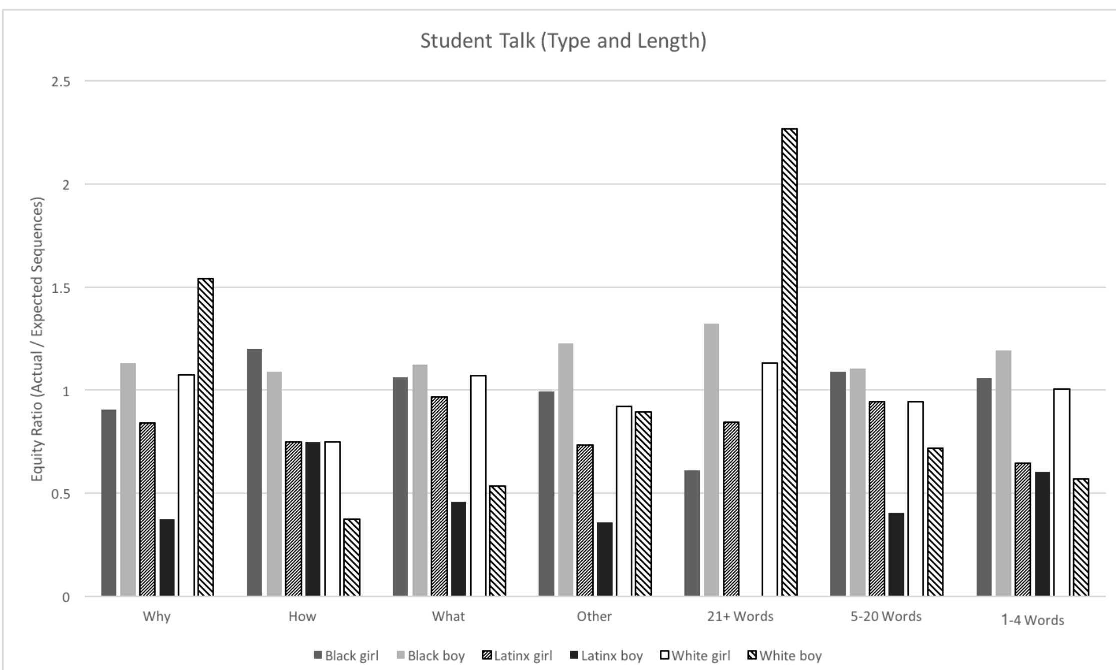

Figure 7. Student Talk (type and length) by race and gender.

Figure 8 shows teacher solicitation types by race and gender. These analytics confirm the patterns of student discourse in Figure 7. Additionally, Figure 8 shows that the small number of spontaneous contributions (the *N/A* category) by Latinx students were from Latinx girls. Fisher's exact test was used to account for small expected values in the contingency table. The results were significant ($p = 0.003$), Cramer's V = 0.124, indicating a small effect size. The standardized adjusted residual for Latinx boys was -3.80, indicating a significant difference; for Latinx girls, it was -2.21, which is borderline (but not quite) significant given the larger contingency table (Sharpe, 2015).

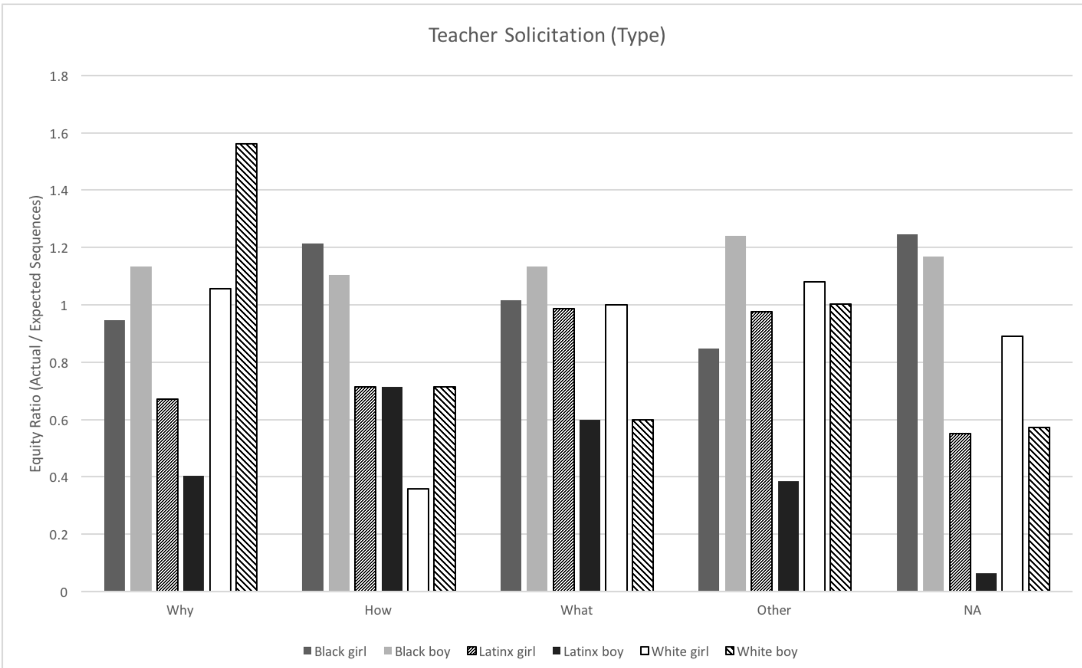

Figure 8. Teacher solicitations (type) by race and gender.

Considering gender alone, it seemed that students had equal opportunities to participate in cognitively demanding ways, but the race-only analytics and intersectional analytics tell an increasingly nuanced story. Although both boys and girls of all races generally had proportional opportunities to participate, it seemed that two groups—White boys and Latinx boys—had differential access in Ms. J's classroom.

Individual Student Results

To visualize individual students, we use a violin plot, which allows us to see the race and gender of individual students alongside dimensions of their participation. Figures 9, 10, and 11 show individual student contributions for Length of Talk, Type of Talk, and Type of Teacher Solicitations, respectively. Note that the axes on these graphs are cut off at 30 to allow for visual comparisons across different levels of the dimensions and that outliers (scores above 30) are indicated in the figure notes. As a result, some outliers in the violin plots are cut off at 30. In these charts, gender is indicated by position relative to the 0-line of the *x*-axis (contributions from boys below and contributions from girls above the 0-line). Racial categories are color-coded. Each violin plot contains an individual half-bar for each of the 30 students in Ms. J's class. Plots with fewer than 30 half-bars indicate that not all 30 students contributed sequences of that type for a given dimension. For example, 5 of the 30 students did not make any contributions longer than 21 words, so the plot for 21+ words in Figure 9 only contains 25 half-bars (instead of 30 half-bars).

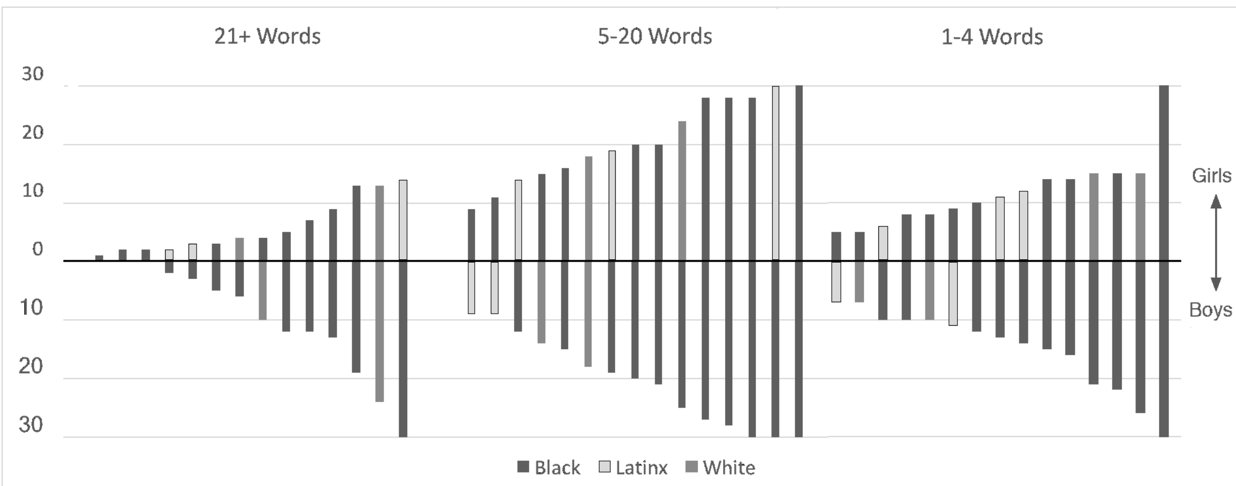

Figure 9. Violin plots of student talk length. Outliers by category. 1–4 words: 70 (Bg), 37 (Bb); 5–20 words: 67 (Bg), 41 (Bb); 21+ words: 37 (Bb), where B indicates Black, g is girl, and b is boy.

Figure 9 shows the violin plots for individual student talk length. Looking at the shapes of the distributions, one can see that extended contributions (21+ words) were limited to a smaller subset of students (as indicated by the narrow tail), whereas many more students participated with shorter contributions (the thicker tails of the distributions for 1–4 words and 5–20 words). Further, the skew of the graphs was larger for boys. This indicates that a few boys who participated frequently with extended contributions were responsible for raising the overall proportion of boys for the 21+ words category. This suggests that differences in extended contributions between boys and girls may have been due to disproportionately greater participation by a few "high-status" boys as opposed to a systematic gender bias across all students. These plots also help further disaggregate the contributions of Latinx students in the class. Surprisingly, it was a Latinx girl—represented by the tall orange bar on the right end of the 21+ words violin plot—who had more extended contributions than any other girl.

Figure 10 shows the distribution of student talk type, and Figure 11 shows the distribution of teacher questions. Analyzing these sets of plots together reveals several interesting insights. For instance, there is a Latinx girl who had the third highest number of *why* statements for a girl, but she was the eighth highest girl for responding to teacher *why* questions. This suggests that this particular student often took up lower level questions as an opportunity for higher level explanation. With the exception of this student, most patterns for the type of student talk generally mimicked patterns for teacher questions. In addition, the distribution of *how* talk and questions shows that many students never explained their reasoning and never responded to these types of questions because they were used very infrequently by the teacher.

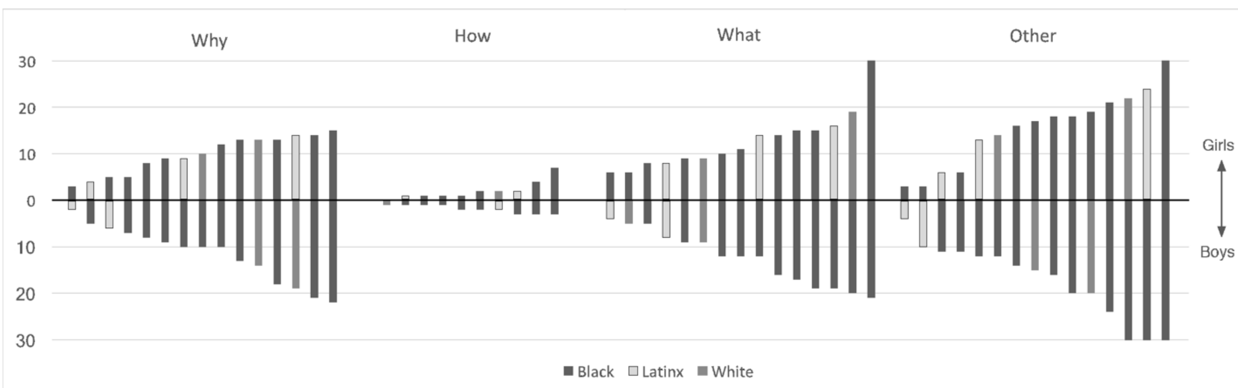

Figure 10. Violin plots of student talk type. Outliers by category of student talk. *Other*: 73 (Bg), 74 (Bb), 36 (Bb), 34 (Bb); *what*: 45 (Bg), where B indicates Black, g is girl, and b is boy.

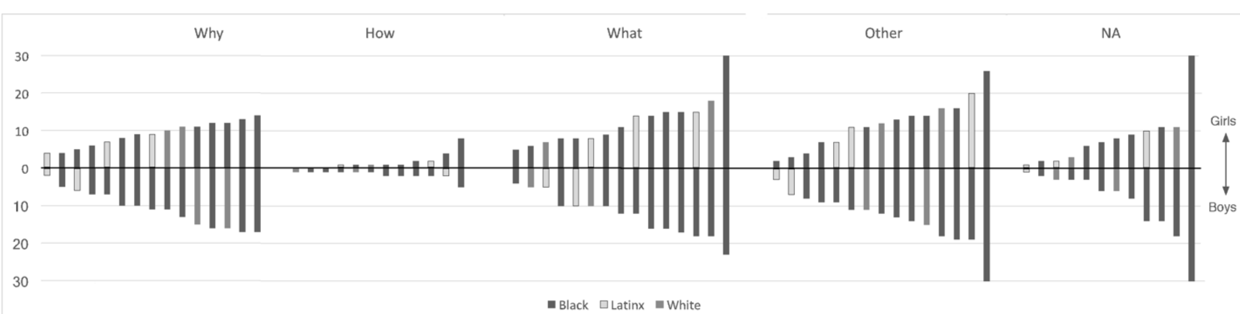

Figure 11. Violin plots of teacher solicitations. Outliers by category of teacher question. *N/A*: 55 (Bg), 33 (Bb); *other*: 45 (Bb); *what*: 36 (Bg), where B indicates Black, g is girl, and b is boy.

In sum, individual-level results provide a way to look at the distribution of student contributions in various categories, which can add a level of nuance beyond group-level results.

Limitations and the Need for Mixed-Methods Approaches

This methodology has limitations. First, EQUIP focuses on lower inference, frequently occurring dimensions of classroom discourse. As a result, EQUIP—and the equity analytics methodology more broadly—may not be as effective at tracking patterns in discursive phenomena that occur infrequently or that require considerable interpretation. For example, in our own prior work, we have found that nonverbal racialized exchanges lasting only a few seconds can have a profound impact on students' mathematical identities and educational experiences (Shah & Leonardo, 2016). Certain higher inference discourse events might be better analyzed by qualitative approaches. Still, the extent to which EQUIP can analyze higher inference phenomena remains an empirical question for future research.

A second limitation relates to the student populations in question. Although EQUIP may provide insights into racial inequities in a racially diverse classroom, in a more homogeneous classroom, it may be less useful for providing insight into the racialized nature of mathematics teaching and learning. Nevertheless, it still may be possible to analyze patterns based on understudied racial phenomena, such as hierarchies based on skin tone (e.g., darker skinned Black students vs. lighter skinned students; Hunter, 2007). Additionally, EQUIP may provide insight into differences among other social markers in such contexts, such as language or gender. In this sense, the value of EQUIP is dependent on the demographic information available and what insights a research team may have into which social markers matter most in a given context.

Third, we acknowledge a potential limitation concerning the size of groups in group-level analyses. In this study, for example, students in Ms. J's class were primarily Black. Although we make claims about White and Latinx boys, there were only two members of each of these groups. As such, one should be cautious in extrapolating strong claims from these groups, given the small sample sizes. At the same time, we note that in classrooms with only a few individuals from a marginalized group, those students are often seen as stand-ins for all members of that group. In other words, even if there are only two Latinx boys in the class, the way they participate may be perceived by their classmates and teacher as representative of how all Latinx boys participate in math classes.

A final limitation is what our approach can actually tell us about equity. Here we argued that equality is a necessary waypoint to equity for students from historically marginalized groups; however, we also stated that equity transcends benchmarks of equality because equity is fundamentally about justice. But what does justice look like in a mathematics classroom? Black students comprised 70% of all students in Ms. J's classroom, and 73% of all high-level questions went to those Black students. Although 73% constitutes a statistically proportional share of high-level questions for Black students, can we say that it is a "just" share? In fact, given the long history of marginalization that Black people have faced in mathematics classrooms and in society, one might argue that equity would involve Black students actually receiving a disproportionately greater share of learning opportunities.

Ultimately, equity is a subjective concept, and the equity analytics methodology cannot—and does not purport to—specify optimal levels and forms of equity. We consider this a limitation intrinsic to all observation tools, which by their nature cannot illuminate subjective facets of equity. It is also critically important that, when possible, studies of equity in classrooms account for students' subjective perceptions of equity. That is, even if the quantitative data indicate more equitable patterns, do the students themselves perceive that they have fair opportunities to participate? Prior research on students' subjective experiences in mathematics classrooms offers a rich base of work from which to build (Berry, 2008; Lubienski, 2000; Martin, 2006; McGee & Martin, 2011; Stinson, 2008).

Tools like EQUIP are most effective when used in conjunction with deeper qualitative analyses that can contextualize and add depth to quantitative data. Further, equity analytics should be interpreted in relation to teachers' professional knowledge of their students and their practice, as well as students' subjective perspectives. In that way, surveys and interviews could be used to complement observational data. These data would provide insight into a teacher's planning and students' subjective experiences, capturing aspects of equity that cannot be ascertained by looking at participation alone. In short, rigid dichotomies between quantitative and qualitative methods are problematic. Integrated methodological approaches are needed to study the multidimensional problem of equity and inequity in classrooms.

Discussion

Classroom discourse is critical to student learning and is thus a site where inequities can proliferate. Indeed, certain groups—non-Asian students of color, girls, and emergent multilingual students, in particular—remain marginalized with respect to their opportunities to participate in discourse (McAfee, 2014; Planas & Gorgorió, 2004; Sadker et al., 2009). This article presented EQUIP as one instantiation of equity analytics, a methodological approach for measuring certain aspects of equity and inequity in discourse. From this perspective, we set out to address this question: What are the affordances and limitations of a quantitative approach to measuring aspects of equity in classroom discourse? In this section, we specify our contribution to the existing literature on measures of classroom phenomena and potential to advance research and practice in the area of equitable instruction.

Ms. J was a highly experienced teacher with deep commitments to equity. In the aggregate, the analytics generated by EQUIP suggest that Ms. J's classroom was an equitable space: Ms. J supported participation in rich sense making through high-level questions and by encouraging student–student discourse. We suspect that existing observation tools, which focus on classroom-level analyses of the quality of classroom discourse, would also support this conclusion. However, because they were not designed for disaggregation at finer grain sizes, existing measures cannot answer *who* in Ms. J's classroom was getting access to these rich participation opportunities.

Equity analytics was designed to address this limitation. In this case, we argue that EQUIP revealed several powerful insights about equity and inequity in Ms. J's classroom that would have been too subtle for other tools to detect. For example, Ms. J distributed participation opportunities proportionally to both girls and Black students. Given that these groups tend to be marginalized in mathematics classrooms (Langer-Osuna, 2011; Martin, 2006; McAfee, 2014), this is a positive finding. It is further noteworthy that Ms. J was able to sustain such patterns over a substantial number of classroom discussions. Nevertheless, without interviewing Ms. J, here we cannot determine why she was able to distribute participation opportunities in this way. Again, this highlights the value of a mixed-method approach.

Alternatively, EQUIP also identified several inequities. With respect to race, White students engaged in a disproportionately greater number of *why* level sequences. Using intersectional analytics to further unpack this pattern, the data show that White boys, in particular, were afforded a disproportionately greater number of *why* level teacher solicitations. Notably, this pattern aligns with what Stinson (2008) has called the "White male math myth" (p. 975), which posits that White men (or boys) inherently possess a greater mathematical ability than others. The analytics also showed that participation by Latinx boys was underrepresented along nearly every dimension. In fact, one of these Latinx boys was Bernardo, the student who appeared in the vignette presented earlier that showed him participating in mathematical discourse. In contrast to the vignette, the individual student-level analytics revealed that this type of participation was, in fact, highly atypical for Bernardo.

We make two notes about these empirical findings. First, it is not surprising that there was evidence of both more and less equitable patterns of interaction in the same classroom. In fact, amidst the density of everyday classroom interaction, we should expect this kind of variation. Observation tools must be sensitive enough to capture both positive and negative equity trends. Second, it also should not be surprising that inequities emerged even in the classroom of a highly skilled, equity-minded educator like Ms. J. The observed inequities might reflect implicit biases that trumped Ms. J's best intentions (Staats et al., 2016). To be sure, it is unlikely that Ms. J realized that she was calling on Latinx students less frequently or that she was consciously trying to privilege the participation of White boys in her class. Without additional data, we cannot make claims about the teacher's intent. However, regardless of intent, the equity analytics illuminate patterns—both promising and problematic—that deserve deeper consideration. These patterns might be explored further in working with the same teacher, or working across a number of different classrooms, to understand which patterns are endemic to our education system more broadly and which are localized to a specific classroom.

We also reiterate that there are tensions associated with achieving equitable participation. For instance, a teacher might use only random calling methods, which would result in an equal distribution of participation across groups. However, using that as the only method could also stifle student agency because students might not necessarily have opportunities to make the contributions that *they* wished to make. Further, it might result in limited student–student cross talk due to the more prescribed nature of interactions. This is not to say that random calling cannot be a useful tool in the service of equity, but it certainly should not be seen as a panacea.

Ultimately, we believe that the real value in the equity analytics methodology lies in its capacity to support educators in noticing subtle equity patterns (Hand, 2012; Wager, 2014).

Because the quantitative data do not speak for themselves, noticing various trends can lead researchers and practitioners to raise generative questions about equitable teaching practice. In the case of Ms. J, what might explain the proportional distribution—and sometimes even overrepresentation—of participation from Black boys in her class? Was this a statistical fluke, or did it reflect a racial justice ideology informing her teaching practice that caused her to deliberately attend to the participation of this group of students? If the latter, then what are the implications for our theorizing of the relationship between equity-related ideologies and teacher practice? Similarly, what might explain the underrepresentation of participation from Latinx boys? Was this an effect of implicit bias, or were these emergent multilingual students whose participation in public discussions Ms. J was purposefully trying to moderate?

Questions of this kind are beyond the scope of the present study but are of central interest in future research. How might equity analytics empower teachers to reflect on equity issues in their own practice? Presumably, if Ms. J had access to such data, she might have made purposeful decisions about how to make her instructional practice more equitable. All teachers need access to actionable data on equity patterns in classroom discourse, especially those related to groups historically marginalized in mathematics education. Compared with existing observation tools that produce holistic descriptions of classrooms, EQUIP is well-positioned to provide such data.

To be clear, in proposing equity analytics we do not mean to perpetuate the narrative of quantitative data or large data sets (i.e., "big data") as having an "aura of truth, objectivity, and accuracy" (d. boyd & Crawford, 2012, p. 663). Data—big or small, quantitative or qualitative—have no inherent meaning. The conclusions that we draw from data are effects of our interpretive frameworks. The goal of equity analytics is not to capture all of the context and rich nuance of classroom interaction. Rather, we argue that certain types of high-leverage quantitative information can provide useful baseline data for exploring fundamental questions about the distribution of participation opportunities.

Conclusion

Classrooms are complex spaces, and inequities can often be too subtle for researchers and practitioners to notice. However, the field currently lacks observation tools that are sensitive enough to illuminate the nuanced ways in which inequity operates in classrooms. This article presented equity analytics—as instantiated in the EQUIP observation tool—as a quantitative approach for capturing certain facets of equity and inequity in classroom discourse. Although our interests and commitments lie primarily in mathematics education, EQUIP's general focus on classroom discourse means that it can also be used in other subject areas. Social forces like patriarchy, racism, linguicism, and other forms of oppression cannot be reduced to a set of statistical comparisons; however, quantitative analytics of the kind described here can complement qualitative approaches that analyze higher inference classroom phenomena consequential for equity. We are cautiously optimistic that the relative simplicity and flexibility of the equity analytics methodology represents a promising step forward for equity work.

References

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, *25*(2), 203–239. doi:10.1080/10508406.2016.1143370

Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2012). Desettling expectations in science education. *Human Development*, *55*(5–6), 302–318. doi:10.1159/000345322

Bartell, T., Wager, A., Edwards, A., Battey, D., Foote, M., & Spencer, J. (2017). Toward a framework for research linking equitable teaching with the standards for mathematical practice. *Journal for Research in Mathematics Education*, *48*(1), 7–21. doi:10.5951/jresematheduc.48.1.0007

Battey, D., Neal, R. A., Leyva, L., & Adams-Wiggins, K. (2016). The interconnectedness of relational and content dimensions of quality instruction: Supportive teacher–student relationships in urban elementary mathematics classrooms. *The Journal of Mathematical Behavior*, *42*, 1–19. doi:10.1016/j.jmathb.2016.01.001

Berry, R. Q., III. (2008). Access to upper-level mathematics: The stories of successful African American middle school boys. *Journal for Research in Mathematics Education*, *39*(5), 464–488.

Bianchini, J. A. (1997). Where knowledge construction, equity, and context intersect: Student learning of science in small groups. *Journal of Research in Science Teaching*, *34*(10), 1039–1065. doi:10.1002/(SICI)1098-2736(199712)34:10<1039::AID-TEA5>3.0.CO;2-S

Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal*, *113*(1), 76–104. doi:10.1086/666387

Boston, M., Bostic, J., Lesseig, K., & Sherman, M. (2015). A comparison of mathematics classroom observation protocols. *Mathematics Teacher Educator*, *3*(2), 154–175. doi:10.5951/mathteaceduc.3.2.0154

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Boyd, M. P., & Rubin, D. L. (2002). Elaborated student talk in an elementary ESoL classroom. *Research in the Teaching of English*, *36*(4), 495–530.

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*(4), 639–669. doi:10.1002/sce.20449

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

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. doi:10.1016/0364-

0213(94)90016-7

Darling-Hammond, L. (1998). Unequal opportunity: Race and education. *The Brookings Review*, *16*(2), 28–32. doi:10.2307/20080779

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press.

Davies, B., & Harré, R. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behaviour*, *20*(1), 43–63.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312. doi:10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

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

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a Community of Learners classroom. *Cognition and Instruction*, *20*(4), 399–483. doi:10.1207/S1532690XCI2004_1

Esmonde, I. (2009). Mathematics learning in groups: Analyzing equity in two cooperative activity structures. *Journal of the Learning Sciences*, *18*(2), 247–284. doi:10.1080/10508400902797958

Esmonde, I., & Langer-Osuna, J. M. (2013). Power in numbers: Student participation in mathematical discussions in heterogeneous spaces. *Journal for Research in Mathematics Education*, *44*(1), 288–315. doi:10.5951/jresematheduc.44.1.0288

Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: A new model for analysis of the educational process. *Educational Research*, *49*(4), 343–363. doi:10.1080/00131880701717198

Foucault, M. (1975). *Discipline and punish: The birth of the prison*. London, UK: Penguin Books.

Gee, J. P. (2011). *An introduction to discourse analysis: Theory and method* (3rd ed.). New York, NY: Routledge.

Goffney, I. D. (2010). *Identifying, measuring, and defining equitable mathematics instruction* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3429420)

Google. (n.d.). *Diversity.* Retrieved from https://www.google.com/diversity/

Green, T. F. (1983). Excellence, equity, and equality. In L. S. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 318–341). New York, NY: Longman.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, *15*(3), 42–57.

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English Language Arts and teachers' value-added scores* (Working Paper No. 16015). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w16015

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. doi:10.1207/S15327833MTL04023_4

Gutiérrez, R. (2007). *Context matters: Equity, success, and the future of mathematics education.*

Paper presented at the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Reno, NV.

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

Hand, V. (2012). Seeing culture and power in mathematical learning: Toward a model of equitable instruction. *Educational Studies in Mathematics*, *80*(1–2), 233–247. doi:10.1007/s10649-012-9387-9

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. doi:10.1080/19312450709336664

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. doi:10.1037/1040-3590.7.3.238

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549. doi:10.2307/749690

Herbel-Eisenmann, B., Choppin, J., Wagner, D., & Pimm, D. (Eds.). (2011). *Equity in discourse for mathematics education: Theories, practices, and policies*. New York, NY: Springer.

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. doi:10.2307/30034933

Hunter, M. (2007). The persistent problem of colorism: Skin tone, status, and inequality.

Sociology Compass, *1*(1), 237–254. doi:10.1111/j.1751-9020.2007.00006.x

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2015). Student participation in elementary mathematics classrooms: The missing link between teacher practices and student achievement? *Educational Studies in Mathematics*, *90*(3), 341–356. doi:10.1007/s10649-015-9625-z

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, *42*(6), 304–316. doi:10.3102/0013189X13495087

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, *27*(1), 29–63. doi:10.3102/00028312027001029

Langer-Osuna, J. M. (2011). How Brianna became bossy and Kofi came out smart: Understanding the trajectories of identity and engagement for two group leaders in a project-based mathematics classroom. *Canadian Journal of Science, Mathematics and Technology Education*, *11*(3), 207–225. doi:10.1080/14926156.2011.595881

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. doi:10.1016/j.tics.2006.08.004

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30–40.

Lubienski, S. T. (2000). Problem solving as a means toward mathematics for all: An exploratory look through a class lens. *Journal for Research in Mathematics Education*, *31*(4), 454–482. doi:10.2307/749653

Martin, D. B. (2003). Hidden assumptions and unaddressed questions in Mathematics for All rhetoric. *The Mathematics Educator*, *13*(2), 7–21.

Martin, D. B. (2006). Mathematics learning and participation as racialized forms of experience: African American parents speak on the struggle for mathematics literacy. *Mathematical Thinking and Learning*, *8*(3), 197–229. doi:10.1207/s15327833mtl0803_2

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.

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

McGee, E. O., & Martin, D. B. (2011). “You would not believe what I have to go through to prove my intellectual value!” Stereotype management among academically successful Black mathematics and engineering students. *American Educational Research Journal*, *48*(6), 1347–1389. doi:10.3102/0002831211423972

Mehan, H. (1979). “What time is it, Denise?”: Asking known information questions in classroom discourse. *Theory Into Practice*, *18*(4), 285–294. doi:10.1080/00405847909542846

Michaels, S., O’Connor, M. C., Hall, M. W., & Resnick, L. B. (2010). *Accountable Talk sourcebook: For classroom conversation that works*. Pittsburgh, PA: University of Pittsburgh.

Milner, H. R., IV. (2010). *Start where you are, but don’t stay there: Understanding diversity, opportunity gaps, and teaching in today’s classrooms*. Cambridge, MA: Harvard Education Press.

Moschkovich, J. N. (2011). How equity concerns lead to attention to mathematical discourse. In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in discourse for* *mathematics education: Theories, practices, and policies* (pp. 89–105). New York, NY: Springer.

Nasir, N. S. (2011). *Racialized identities: Race and achievement among African American youth*. Stanford, CA: Stanford University Press.

Nasir, N. S., & Shah, N. (2011). On defense: African American males making sense of racialized narratives in mathematics education. *Journal of African American Males in Education*, *2*(1), 24–45.

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.* Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Author.

National Research Council. (2000). *How people learn*: *Brain, mind, experience, and school.* Washington, DC: National Academy Press.

National Research Council. (2013). *Adapting to a changing world–Challenges and opportunities in undergraduate physics education*. Washington, DC: National Academies Press.

Neumann, M. D. (2014). Preservice teachers' understanding of gender equity in K-6 mathematics teaching. *Teacher Education and Practice*, *27*(1), 90–117.

Oakes, J. (2005). Keeping track: How schools structure inequality. New Haven, CT: Yale University Press.

Orfield, G., & Lee, C. (2007). *Historic reversals, accelerating resegregation, and the need for new integration strategies*. Retrieved from http://files.eric.ed.gov/fulltext/ED500611.pdf

Parks, A. N. (2010). Explicit versus implicit questioning: Inviting all children to think mathematically. *Teachers College Record*, *112*(7), 1871–1896.

Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, *18*(2), 133–145. doi:10.2307/3150948

Planas, N., & Gorgorió, N. (2004). Are different students expected to learn norms differently in the mathematics classroom? *Mathematics Education Research Journal*, *16*(1), 19–40. doi:10.1007/BF03217389

President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the president.* Washington, DC: Executive Office of the President.

Resnick, L. B., Michaels, S., & O'Connor, M. C. (2010). How (well-structured) talk builds the mind. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 163–194). New York, NY: Springer.

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, *37*(1), 43–50. doi:10.1177/002248718603700110

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.

Samoff, J. (1996). Which priorities and strategies for education? *International Journal of Educational Development*, *16*(3), 249–271. doi:10.1016/0738-0593(96)00017-X

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist*, *23*(2), 145–166.

doi:10.1207/s15326985ep2302_5

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational Researcher*, *43*(8), 404–412. doi:10.3102/0013189X14554450

Schoenfeld, A. H. & the Teaching for Robust Understanding Project. (2016). An introduction to the Teaching for Robust Understanding (TRU) framework. Berkeley, CA: Graduate School of Education. Retrieved from http://map.mathshell.org/trumath/intro_to_tru_20161223.pdf

Schultz, K. (2009). *Rethinking classroom participation: Listening to silent voices*. New York, NY: Teachers College Press.

Secada, W. G. (1989). Educational equity versus equality of education: An alternative conception. In W. G. Secada (Ed.), *Equity and education* (pp. 68–88). New York, NY: Falmer.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. New York, NY: Cambridge University Press.

Shah, N. (2017). Race, ideology, and academic ability: A relational analysis of racial narratives in mathematics. *Teachers College Record*, *119*(7).

Shah, N., & Leonardo, Z. (2016). Learning discourses of race and mathematics in classroom interaction: A poststructural perspective. In I. Esmonde & A. N. Booker (Eds.), *Power and privilege in the learning sciences: Critical and sociocultural theories of learning* (pp. 50–69). New York, NY: Routledge.

Shah, N., & Lewis, C. M. (in press). *Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework.* [Journal name].

Sharpe, D. (2015). Your chi-square test is statistically significant: Now what? *Practical Assessment, Research & Evaluation*, *20*(8). Retrieved from http://pareonline.net/getvn.asp?v=20&n=8

Soter, A. O., Wilkinson, I. A., Murphy, P. K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research*, *47*(6), 372–391. doi:10.1016/j.ijer.2009.01.001

Staats, C., Capatosto, K., Wright, R. A., & Jackson, V. W. (2016). *State of the science: Implicit bias review*. Columbus, OH: Kirwan Institute. Retrieved from http://kirwaninstitute.osu.edu/wp-content/uploads/2016/07/implicit-bias-2016.pdf

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

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488. doi:10.3102/00028312033002455

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. doi:10.3102/0002831208319723

Tanner, K. D. (2013). Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE—Life Sciences Education*, *12*(3), 322–

331. doi:10.1187/cbe.13-06-0115

Turner, E., Dominguez, H., Maldonado, L., & Empson, S. (2013). English learners' participation in mathematical discussion: Shifting positionings and dynamic identities. *Journal for Research in Mathematics Education*, *44*(1), 199–234. doi:10.5951/jresematheduc.44.1.0199

University of Michigan. (2006). *A coding rubric for measuring the mathematics quality of instruction* (Technical Report LMT1.06). Ann Arbor, MI: University of Michigan, School of Education.

Wager, A. A. (2014). Noticing children's participation: Insights into teacher positionality toward equitable mathematics pedagogy. *Journal for Research in Mathematics Education*, *45*(3), 312–350. doi:10.5951/jresematheduc.45.3.0312

Wagner, D., Herbel-Eisenmann, B., & Choppin, J. (2011). Inherent connections between discourse and equity in mathematics classrooms. In B. Herbel-Eisenmann, J. Choppin, D. Wagner, & D. Pimm (Eds.), *Equity in discourse for mathematics education: Theories, practices, and policies* (pp. 1–13). New York, NY: Springer.

Weber, K., Radu, I., Mueller, M., Powell, A., & Maher, C. (2010). Expanding participation in problem solving in a diverse middle school mathematics classroom. *Mathematics Education Research Journal*, *22*(1), 91–118. doi:10.1007/BF03217560

Wood, M. B. (2013). Mathematical micro-identities: Moment-to-moment positioning and learning in a fourth-grade classroom. *Journal for Research in Mathematics Education*, *44*(5), 775–808. doi:10.5951/jresematheduc.44.5.0775

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458–477.

doi:10.2307/749877

Authors
Daniel L. Reinholz, Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-7720; daniel.reinholz@sdsu.edu
Niral Shah, Department of Teacher Education, Michigan State University, 116P Erickson Hall, East Lansing, MI 48824; niral@msu.edu

Submitted January 23, 2017
Accepted June 21, 2017

APPENDIX A

TRU Math Dimensions

The Five Dimensions of Powerful Classrooms				
The Content	**Cognitive Demand**	**Equitable Access to Content**	**Agency, Ownership, and Identity**	**Formative Assessment**
The extent to which classroom activity structures provide opportunities for students to become knowledgeable, flexible, and resourceful disciplinary thinkers. Discussions are focused and coherent, providing opportunities to learn disciplinary ideas, techniques, and perspectives, make connections, and develop productive disciplinary habits of mind.	*The extent to which students have opportunities to grapple with and make sense of important disciplinary ideas and their use. Students learn best when they are challenged in ways that provide room and support for growth, with task difficulty ranging from moderate to demanding. The level of challenge should be conducive to what has been called "productive struggle."*	*The extent to which classroom activity structures invite and support the active engagement of all of the students in the classroom with the core disciplinary content being addressed by the class. Classrooms in which a small number of students get most of the "air time" are not equitable, no matter how rich the content: all students need to be involved in meaningful ways.*	*The extent to which students are provided opportunities to "walk the walk and talk the talk" – to contribute to conversations about disciplinary ideas, to build on others' ideas and have others build on theirs – in ways that contribute to their development of agency (the willingness to engage), their ownership over the content, and the development of positive identities as thinkers and learners.*	*The extent to which classroom activities elicit student thinking and subsequent interactions respond to those ideas, building on productive beginnings and addressing emerging misunderstandings. Powerful instruction "meets students where they are" and gives them opportunities to deepen their understandings.*

Schoenfeld, A. H., & the Teaching for Robust Understanding Project. (2016). An Introduction to the Teaching for Robust Understanding (TRU) Framework. Berkeley, CA: Graduate School of Education. Retrieved from http://map.mathshell.org/trumath.php.

APPENDIX B

Excerpt from Extension of MQI: Mathematics Quality and Equity (MQE)

V. Use of mathematics to teach equitably (measured for each clip)

Dimension	**Present / Not Present**
a1. Real-world problems or examples not present	
a2. Real-world problems or examples present	
b. Explicit student tasks and work	
c. Explicit talk about the meaning and use of mathematical language	
d. Explicit talk about ways of reasoning	
e. Explicit talk about mathematical practices	
f. Instructional time is spent on mathematics	
g. Teacher encourages diverse array of mathematical competence	
h. Teacher emphasizes student effort and conveys message that effort will eventually pay off	
i. Teacher encourages and gives opportunities for students to work autonomously	
j. Expressed expectation that everyone will be able to do the work	

(excerpt from Goffney, 2010)

APPENDIX C

Excerpt from Accountable Table Rubrics (Instructional Quality Assessment)

I. How effectively did the lesson-talk build Accountability to the Learning Community?

Participation in the Learning Community
Was there widespread participation in teacher-facilitated discussion?

Rubric 1: Participation	
4	Over 75% of the students participated throughout the discussion.
3	50-75% of the students participated in the discussion.
2	25-50% of the students participated in the discussion.
1	Less than 25% of the students participated in the discussion.
0	None of the students participated in the discussion.
N/A	Reason:

_______ Number of students in class

_______ Number of students who participated

(from:
http://peabody.vanderbilt.edu/docs/pdf/tl/IQA_RaterPacket_LessonObservations_Fall_12.pdf)