SN Soc Sci (2023) 3:76
https://doi.org/10.1007/s43545-023-00664-0

ORIGINAL ARTICLE

Visualizing inequity: how STEM educators interpret data visualizations to make judgments about racial inequity

Daniel L. Reinholz[1] · **Samantha Ridgway**[2] · **Poorna Talkad Sukumar**[3] · **Niral Shah**[4]

Received: 5 October 2022 / Accepted: 21 April 2023

Abstract

Data visualizations are routinely used for STEM faculty development to support equitable teaching practices. Yet, little is known about how instructors interpret such data visualizations. This interview study fills a key gap by providing insight into how STEM educators make sense of visualizations. We report on cognitive interviews with 17 participants who were shown eight different data visualizations depicting racial inequities in classroom participation. The participants were asked to interpret whether the scenarios were equitable and answer questions about the distribution of participation. We report on which visualizations participants were able to interpret most accurately, and how particular visualizations supported thinking about inequity. No single visualization was most effective in all cases, and critically, we found that not all visualizations were equally effective for identifying inequities, and that different types of visualizations drew attention to different aspects of inequity (e.g., individual disparities vs. group-level disparities). We also provide data on how participants differentiated between equity and equality. Thus, the present study provides useful information for professional developers about which types of visualizations may be most effective for different purposes and highlights the need for multiple representations of racial inequity.

Keywords Data analytics · Equity · Professional development · Race

✉ Daniel L. Reinholz
daniel.reinholz@sdsu.edu

1 Department of Mathematics & Statistics, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-7720, USA

2 Center for Research in Mathematics and Science Education, San Diego State University, San Diego, CA 92120-5013, USA

3 Department of Technology Management and Innovation, NYU Tandon School of Engineering, 6 MetroTech Center, Brooklyn, NY 11201, USA

4 Learning Sciences and Human Development, University of Washington, Seattle, WA 98195-3600, USA

Published online: 05 May 2023

Data are routinely used to support professional development in STEM higher education. Data come in many forms, including student scores on concept inventories (Marbach-Ad et al. 2010), results from classrooms observation tools (Smith et al. 2013), or even personas (Zagallo et al. 2019). Recently, work has emphasized the role of data visualizations to support deeper thinking about racial equity in teaching (Reinholz et al. 2020a, b; e.g., Reinholz et al. 2022; Reinholz and Shah 2021; Shah et al. 2020). Although such data have shown promise for professional development, less is known in specific about how instructors make sense of equity-focused data. However, a deep understanding of how instructors interpret such data is a necessary step towards developing more effective visualizations, which can further catalyze more effective professional development organized around data. This is the central focus of the current study.

The present study consisted of cognitive interviews with 17 participants (13 STEM educators and 4 American Sign Language [ASL] interpreters for STEM classes). The ASL interpreters were considered important to include, because far too often, the concerns of disabled students are left out of efforts to improve equitable teaching. These 17 participants were asked the same set of interview questions for eight different visualizations describing scenarios of racial in/equity (described in the methods). Analyses of participant responses helped answer the following sets of research questions:

RQ1. Which data visualizations supported participants to correctly interpreting aspects of student participation?
RQ2. What visualization preferences did participants have and which features of the visualizations were most helpful?
RQ3. How did these visualizations mediate participant reasoning about equity?

By answering these research questions, this manuscript makes an important contribution to the research literature as it provides a basis for understanding how STEM educators interpret data analytics, which is necessary for building more effective data analytics for professional development. Particularly, this manuscript can help push the field forward to develop more effective methods for professional development around racial equity. Given that our sample draws from postsecondary STEM educators, our results have direct implications for STEM faculty development.

Theoretical framing

Equity analytics

We use *equity analytics* as our broad analytical approach (Reinholz and Shah 2018), which aims to answer the question: to what extent does the actual distribution of resources in an educational system align or diverge from the distribution predicted based on demographic representation? Possible resources for learning are varied, from qualified teachers (Oakes 2005) to culturally relevant curricula and pedagogy (Ladson-Billings 1995). For the present study, we focus specifically on participation as a resource for learning, or *participatory equity*, which concerns the fair distribution of participation and opportunities to participate in the classroom (Shah and Lewis 2019). Participation is an important resource for learning, as it helps students learn content (Banes et al. 2018; Ing et al. 2015) and develop disciplinary identities (Nasir 2002). Forms of meaningful participation are varied, including both verbal and non-verbal interactions (e.g., gesture, taking notes). Thus, when considering whether participatory equity has been achieved, it is important to consider the quantity, quality, and forms of participation.

While equity can be difficult to define, it can be even harder to operationalize. What would a "fair" distribution of student participation look like? As external observers, it would be problematic for us to tell students what they need, and certainly, any concept of fairness should account for whether students subjectively *feel* that something is fair. At the same time, one should be able to identify glaring inequities in the education system. Going one step further, could equity be operationalized by productively visualizing it? Data alone are insufficient; they must be presented in a way that facilitates effective thinking.

One tool is to consider equity as a waypoint to equality (Secada 1989). Equality is simple to define and measure. It describes a situation where all students would get "the same thing." In terms of participation, it means that all students have the same amount and quality of participation. Equality ignores race, culture, gender, ability, and any other form of difference, and as such, it can never result in true equity. To be clear, we argue that equity and equality are distinct concepts and are not the same, and it can be problematic when these two ideas are conflated. Nonetheless, we argue that, especially for students from minoritized populations, we can think of equality as a necessary, but insufficient baseline. From our viewpoint, achieving equity must account for historical legacies of discrimination and oppression (Darling-Hammond 1998; Tate 2008). This means that for minoritized students, receiving less than equality is clearly an inequity, but realistically, they may need to receive *more* than an equal share of participation opportunities, to account for legacies of inequity. From this viewpoint, one could argue that minoritized students should receive *more* than a proportional share of opportunities, a so-called reparations stance. Still, it may be impossible to identify a specific allocation of resources or participation opportunities that would signify equity has "been achieved." In practice, however, we find that in most of today's classrooms, the amount of actual participation and participation opportunities for minoritized students are well below equality (Ernest et al. 2019; Reinholz et al. 2022a, b; Reinholz and Wilhelm 2022; Shah et al. 2020), signifying the considerable work to be done.

In prior studies, we have operationalized the principles of equity analytics through the EQUIP observation tool (Reinholz and Shah 2018). EQUIP provides a methodology for tracking patterns of verbal student participation in classrooms, broken down by different demographic groups.[1] Typically, an observer uses the EQUIP protocol to code the features of student participation during a lesson (either a live

[1] We have also used EQUIP to capture nonverbal participation, which will be published in forthcoming work.

observation or video recording), which are combined with student demographics to generate analytics. Thus, the EQUIP tool makes it easy to provide actionable data to instructors to support professional development (Reinholz et al. 2020a, b). This work involves providing data visualizations to support cycles of reflection and professional development, through which instructors iteratively revise their teaching practices to address inequities found in their classroom data. These data have the potential to allow instructors to better see their own implicit biases, and consequently, work to mitigate them through intentional strategies. At the same time, field studies with EQUIP highlights that all visualizations are not equally effective. In fact, some visualizations may even promote thinking that works *against* classroom equity (Reinholz and Shah 2021). This work motivated the present study, as it highlights the pressing need to identify more effective data visualizations to promote professional development. The present study complements prior field studies because it allows us to study data visualizations in a more controlled context.

Visualization design

Data visualization has proven to be a powerful technique to facilitate analytic reasoning and improve decision-making in various domains. Data visualization also has the potential to provide important insights to STEM educational researchers. Yet, this body of work has rarely been used to support better visualization design to promote more effective professional development. Data visualizations can help make subtle phenomena, like bias, easier to see by helping individuals "surpass the limitations of [their] own internal cognition" (Munzner 2014). Properly designed visualizations can help mitigate cognitive biases (Dimara et al. 2017; Valdez et al. 2018), and this approach has been proposed to reduce inequities in the college admissions process (Sukumar et al. 2018; Sukumar and Metoyer 2018). This work strongly suggests that insights from data visualizations can also contribute productively to the reduction of implicit biases in STEM teaching, but this hypothesis needs to be further explored.

The methodologies used in data visualization studies broadly have the potential to support education research (Lam et al. 2012). Visualization solutions of classroom participation data can be generated using existing visualization design principles and guidelines (Munzner 2014). Likewise, solutions can be rigorously evaluated using various methods to determine which among them are most effective in leading teachers to recognize their biases and to make more equitable decisions. The methodologies can also unveil what makes the visualizations effective. For example, comparative evaluations of the visualization solutions can be conducted using example tasks (e.g., asking teachers to identify differences in the participation of two populations) and using measures, such as task accuracy and mental effort (Saket et al. 2019). Further, qualitative methods, such as cognitive interviews, can be useful to gather subjective preferences and the "why" and "how" aspects concerning the visualization solutions (Lam et al. 2012). Coalescing findings from such studies can not only enable us to determine how visualizations promote reasoning in the domain of participatory inequity but can also inform visualizations of demographic

Table 1 Participant demographics ($N = 17$ total)

	Asian	Black	Chicanx	White	Biracial*	Total
Woman	1	1		10	2	14
Man			1	1		2
Nonbinary (she/they)			1			1
Total	1	1	1	11	2	17

*One participant was White/Chinese, and the other was Persian/Chinese

**14 participants identified as bilingual; 3 participants identified as having a disability

***The average age of the sample was 37.6 (SD = 8.3) years

inequities in other similar domains, such as selection and hiring processes, and contribute to the knowledge of mitigating implicit biases using visualizations in general.

Methods

Participants

A total of 17 participants ($N = 10$ Tenure-Track STEM Faculty, $N = 3$ STEM Education Doctoral Candidates, $N = 4$ ASL Interpreters for STEM courses, all full-time employees) were recruited through social media and personal communications. This sample size is generally considered sufficient for the purpose of a comparative useability study, such as ours (Macefield 2009). Moreover, because our primary goal was support theoretical generalization about participant sensemaking (Yin 2009), we were more concerned with the underlying reasons behind *why* particular visualizations were effective or not, rather than attempting to generalize statistically with large population.

Participants were not provided any incentives for participation. A variety of demographic information was collected from participants. Overall, our sample consisted of more women than an average sample of STEM faculty (National Academy of Sciences 2007). This reflects the fact that education fields tend to be more skewed towards women, and several STEM faculty members in our sample had interests in education. In this way, our sample was more closely aligned with K12 STEM teachers, who are majority white women (Nguyen and Redding 2018) (Table 1).

Data visualization design

The interviews were organized around eight data visualization scenarios (See Table 2). Seven of the graphs were variations of bar charts, while the eighth was a dot array visualization. We, however, group the visualizations based on the data they present and not based on their types. We categorize the eight visualizations into

Table 2 The visualizations in the eight scenarios

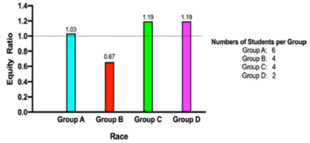
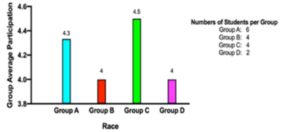
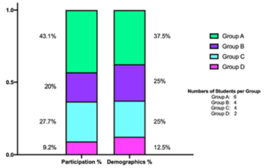
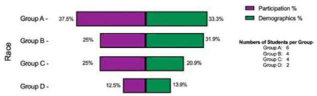
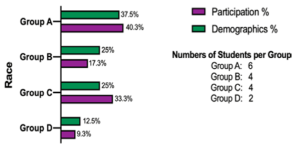
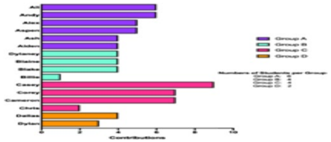
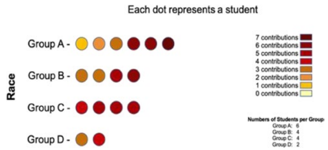
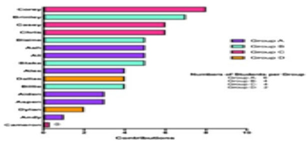
Name/type	Visualization	Description/examples
Equity ratio (derived)		Vertical bar chart. *Equity Ratio = Actual Participation/Expected Participation* Suppose 4/16 students are Latinx, with 24/48 contributions, then actual participation is 50% (24/48), whereas demographic representation is only 25% (4/16), so the equity ratio would be 2 (50/25). Ratio > 1, indicates that Latinx students participated more than expected based on demographic representation
Group average participation (derived)		*Group Average Participation = Actual Participation/Group Size* In the above example, 24 contributions/4 Latinx students = 6. Computationally simpler than an equity ratio
Stacked bars (group)		Two vertical stacked bar charts side-by-side show the breakdown of proportions of participation (left) and classroom demographic representation (right)
Mirror bars (group)		Two horizontal bar charts mirrored side-by-side show the breakdown of proportions of participation (left) and classroom demographic representation (right)
Paired bars (group)		Four sets of paired horizontal bar charts show the breakdown of proportions of participation (top) and classroom demographic representation (bottom)

Table 2 (continued)

Name/type	Visualization	Description/examples
Demographic bars (individual)		Horizontal bar chart showing the number of contributions from each student. An individual student's racial group is represented by the color of their bar. Bars are ordered by group, and each group is ordered from most contributions to least
Dot array (individual)		The dot array shows the amount of participation from each student based on the color intensity of each dot. Darker shades of red indicate more participation
Ordered bars (individual)		Horizontal bar chart showing the number of contributions from each student. An individual student's racial group is represented by the color of their bar. Bars are ordered from most contributions to least, irrespective of group membership

three groups: (1) Derived Data graphs, (2) Individual graphs, and (3) Group graphs. We generally followed common language from the field of data visualizations for the names of the graphs in Table 2, but when no standard term existed, we created our own term for easy reference throughout the manuscript (e.g., Demographic Bars).

For this study, we focused only on data pertaining to the *quantity* of student participation rather than the *quality* of participation. The rationale behind this decision was that there were already numerous visualization scenarios, so we aimed to simplify the actual data presented. Moreover, we see a focus on equitable *quantities* of participation an important first step to address inequitable *qualities* of participation. Future work would instead focus on both quantity *and* quality, with a fewer number of scenarios.

These eight visualizations were chosen based on prior work with the EQUIP tool (Reinholz et al. 2020a; Reinholz and Shah 2018), and a preliminary exploration of a design space of possible solutions for visualizing classroom participation data (P. Sukumar et al. 2020). This was grounded in visualizations literature which describes best practice for design; for example, bar charts are a highly preferable and effective type of graph for most tasks, compared to an alternative such as line chart or pie chart (Saket et al. 2019). The chosen visualizations represent a set of *valid* solutions (Munzner 2014) in terms of the visual encodings given the data types and the tasks that teachers are expected to perform with the visualizations to interpret the data and reflect on their biases. The participants were told that "groups" in the visualizations represented different racial groups to spur sensemaking about racial inequity. However, they were not given any specific races to avoid the possible confounding variable of biases against particular races (e.g., anti-Blackness in STEM; Martin 2019). The number of students in each racial group was held constant to reduce cognitive load (Group A = 6 students, Group B = Group C = 4 students, Group D = 2 students), and this information was listed on each slide. The datasets used in the visualizations were randomly generated using a normal distribution centered at four contributions per student (16 students total), with a standard deviation of 2. Then an absolute value was applied to remove any negative values. Based on our experience from prior field studies (e.g., Reinholz et al. 2020a, b; Reinholz and Shah 2018; Shah et al. 2020), the randomly generated datasets provided a fair emulation of real classroom scenarios that represented some racial inequity. This means that inequities between groups were visible, but not so glaring as to obviate the use of analytics to enhance reasoning.

The visualizations were generated using GraphPad Prism software.[2] We used static visualizations to emulate what could be given to a teacher in a static, paper-based report. In future work, we will explore the role of interactive visualizations. The order of these visualizations was randomized between participants to account for any ordering effects. Between each participant, visualizations were rotated in groups based on the style of graph (*Derived Data*, *Individual*, *Group*) and each graph within the group was rotated so the graphs within a particular group had a mixed order. With each visualization, participants were shown the size of the four

[2] The dot array graph was generated using both GraphPad Prism and Microsoft Powerpoint.

Table 3 Interview segments and research questions

Segments	Research questions answered
Segment 1: Questions	
1. What does the term equity mean to you?	RQ3
2. How do you think about equity in terms of teaching in a classroom?	–
3. Some people make a distinction between equity and equality. In what ways do you think these two concepts are the same or different?	–
Segment 2: Questions (same questions for each of 8 visualizations)	
1. In your opinion, is the distribution of student participation equitable? Explain why or why not	RQ3
2. Which group of students participated the most, overall (i.e., which group had the highest overall number of contributions?)	RQ1
3. Which group of students had the highest average participation per student (i.e., in which group did students contribute the most, relative to the size of the group)?	RQ1
4. Suppose Amara also belongs to Group A. If you assume that Amara participated like an average student in Group A, how many times did they participate?	RQ1
Segment 3: Questions	
1. Which graphs did you find the easiest to interpret? (Rank order the top 3 graphs)	RQ2
2. Why did you choose these particular graphs? What was helpful about them?	RQ2
3. What information did you feel like was missing from the graphs?	RQ2
4. Was there anything you found difficult or confusing?	RQ2
5. Do you have any new thoughts about equity after completing this study?	–
6. Is there anything else we should know?	–

groups that represented racial groups (without any labels such as Black or Latinx). The data in each visualization were explained to participants, and if there was a computed value in the visualization, the computation was provided.

Interview structure

Each interview was conducted over Zoom, for approximately one hour. All interviews were conducted by the second author. The interviews comprised three segments, which together answered our three research questions: (1) general conceptions of equity, (2) eight visualizations/charts with classroom participation scenarios, and (3) debrief questions. The interview questions asked in the three segments and the research questions they correspond to are presented in Table 3.

The purpose of the questions in Segment 1 was to collect general information about participant's conceptions and to prepare them to answer more detailed equity questions with respect to the graphs. The second segment focused on the interpretation of graphs of student participation. Classroom participation was defined for participants using the following language,

> The following graphs show how often different students in a classroom *contributed* to a discussion verbally. A contribution by a student consists of a few words to multiple sentences that were shared without interruption from any other students.

With each visualization, the participants were asked a set of four interview questions that were designed to represent the real types of interpretations that teachers would need to make if they were using the data visualizations to reflect on their own teaching. Participants were told that it may not be possible to answer all questions with all visualizations, so “not possible” was also a valid answer. All visualizations provided the information to answer all questions except some could not unambiguously answer number 4, because these graphs visualized percentages and not absolute values for the contributions (in particular, Equity Ratio, Stacked Bars, Mirror Bars, and Paired Bars). These visualizations are excluded from the analysis below for that question.

The third segment of the interview consisted of debrief questions. These questions were used to get further depth about which visualizations participants found useful and why. If needed or requested by the participants, they were shown all the different visualizations used in the study to support recall.

Participants were allowed an unlimited amount of time to answer a particular question, and the interviews only moved forward when they were ready. If participants had questions at any time during the interview, they were encouraged to ask for clarification, which was provided. In addition, participants were prompted to articulate their thinking verbally, following a think-aloud protocol for cognitive interviews (Beatty and Willis 2007). This allowed us to determine both when responses were correct and how participants were thinking about the visualizations. All interviews were then transcribed for analysis.

Analysis

To understand participant responses, we engaged in a variety of qualitative and quantitative analyses. Our first analysis focused on RQ1, that is, whether participants could correctly answer interview questions 2–4 from Segment 2, which concerned interpretation of the visualizations and the distribution of student participation. For interview questions 2 and 3, we were interested to see if participants could identify the groups(s) that had the greatest overall participation, and greatest proportion of participation per student in each group, respectively. In interview question 4, participants were asked to compute an average amount of participation for Amara. Because participants were not using a calculator, we were interested to see if they could get a relatively accurate sense of participation. Moreover, because fractional participation doesn’t truly make sense in a classroom, we coded all answers with an error tolerance or range of one. Thus, if a participant estimated 3.2, we would consider this correct if the true value was within the closed interval [2.7, 3.7]. If a participant provided an interval of length one, such as “between 3 and 4,” the answer was coded as correct if it was in the interval [3,4].

Our coding scheme had three levels. We coded a response as a *correct answer* if the participant chose the groups that answered the question (questions 2 and 3), or if they provided a correct numerical answer. We coded a response as a *correct strategy* if the participant provided a strategy that, if executed properly, would result in the correct answer. These categories were mutually exclusive; if a participant provided a strategy and answer, it would be counted as *correct answer*. We consider correct answers to be more useful than correct strategies, even though both are technically correct. In prior field studies, we have found that participants tend to use their first impressions of the data visualizations, without putting in the effort to make extended computations (i.e., correct strategies are less reflective of usage in real life). If a participant did not provide either of these, it was marked as *incorrect*. We also tabulated participant responses to the question of which visualizations they found most effective (from Segment 3), to allow comparisons between performance and preference.

Participants were not allowed partial correctness, because we were interested in the absolute accuracy of responses. Similarly, if there were multiple correct answers, they had to identify all of them to be correct. The visualization that had multiple potential correct answers for Question 2 was Ordered Bars, and the visualization that had multiple correct answers for Question 3 was Equity Ratio (both by random chance). Because these two visualizations differed from the rest in the number of correct answers, we elaborate participant responses to these visualizations in the results section.

To understand participants' processes of reasoning and their subjective experiences of working with the visualizations (RQ2), we engaged in an iterative constant thematic analysis (Glaser and Strauss 1967). We used MaxQDA® to support our analyses. The first step of our process involved breaking up the transcripts by question (given in Table 2). This allowed us to code all responses for each graph at the same time, without focusing on which participants provided which responses (i.e., we could avoid any biases based on our perceptions of particular participants).

Our first round of actual coding focused on identifying all instances in which something "helpful" about a graph was noted. A total of 37 instances were identified. Each instance consisted of a single statement about a particular graph or group of graphs. If multiple features were described, multiple instances were coded. Next, we iteratively developed themes associated with each of these instances. The main themes were *seeing individuals* ($N=12$), *calculations provided* ($N=10$), *and side-by-side comparisons* ($N=7$). There were four other codes that came up infrequently and did not hold up as themes, which are described in the results.

The second round of coding identified all instances in which something "missing" about a graph was noted. A total of 26 instances were coded in which participants discussed something that was missing or that they wanted to know more about. We found 11 themes, which are described in the results in Table 7.

The third round of coding focused on how participants made sense of equity (RQ3). First, we considered participants' general conceptions (from Segment 1). We found these broke down into three categories: *equity as equality* ($N=3$), *equity requires appropriate support* ($N=13$), or a *technical definition* ($N=2$). Next, we broke down the determinations of equity by visualization type (question 1 from Segment 2), to see if there could be mediating factors in the visualizations themselves.

Table 4 Performance on question 2, which group participated most overall?

Type	Visualization	% Correct (answer provided)	% Correct (only strategy given)	% Correct (by type)
Derived data	Equity ratio	23.5	0	32.35
	Group avg participation	41.2	0	
Group	Stacked bars	94.1	0	94.1
	Mirror bars	100	0	
	Paired bars	82.4	5.9	
Individuals	Demographic bars	56.3	6.3	
	Dot array	75	0	47.8
	Ordered bars	5.9	0	

Here, we drew out general themes across the different types of visualizations (i.e., *Individual, Group,* and *Derived Data*).

Results and discussion

Visualization performance (research question 1)

In this section, we break up participant performance for each interview question for each visualization type. Interview question 2 focused on which visualizations supported thinking about group *total* participation (see Table 4). The *Derived Data* graphs resulted in relatively low performance. The computations did not directly provide information about which groups participated the most overall—this information had to be computed from the participation rates combined with the size of the groups. This was not intuitive, especially for the equity ratio, because raw percentages were obscured by the calculations used to generate the visualizations.

Participants had the most correct responses when they used visualizations from the *Group* category, as we had predicted. The visualizations in this category were well-suited to the task at hand, because the proportion of participation of each group could be read directly from the graphs, and these proportions corresponded directly to the total quantity of participation.

The *Individual* graphs had mixed performance. The Demographic Bars and Dot Array were relatively effective for determining which group participated most overall, because the individual students were sorted by demographic group. Ordered Bars was not as effective, because individual students were mixed up across groups, so they were hard to aggregate into groups. Thus, the patterns of correct responses generally reflected our theoretical predictions.

The next performance interview question focused on average participation within a group (Question 3, see Table 5). In this case, we hypothesized that the *Derived Data* graphs provided an advantage because they directly computed the ratios. Yet, surprisingly, the percentage of correct answers was only 52.55%. Although the Equity Ratio graph trivially provided the answer, only 23.5% of participants could

Table 5 Performance on Question 3, which group had the highest average participation?

Type	Visualization	% Correct (answer given)	% Correct (strategy only)	% Correct (by type)
Derived data	Equity ratio	23.5	0	43
	Group avg participation	62.5	0	
Group	Stacked bars	52.9	5.9	56.9
	Mirror bars	17.7	17.7	
	Paired bars	64.7	11.8	
Individuals	Demographic bars	64.7	5.9	60.3
	Dot array	68.8	6.3	
	Ordered bars	35.3	0	

correctly identify the answer with this graph. Six participants said that it was not possible to answer, and five participants only identified one of the two groups that had the highest percentage of participation. Given that these visualizations can trivially answer the interview questions if someone correctly interprets the graph, one would have expected better performance. This suggests that participants did not fully understand the visualizations, or perhaps the question. For example, in some cases participants appeared to incorporate the sizes of the student groups to make their determination, which in this case, was irrelevant information. In a field setting, participants would need extra support to fully understand the *Derived Data* visualizations in an intervention setting.

The *Group* graphs had mixed performance. These graphs did not directly provide the required information, so participants needed to use a ratio of percentage of contributions and percentage of demographic representation. Of the *Group* graphs, the Paired Bars made this easiest to do, because the demographic and participation percentages were placed adjacently, which made them easy to compare. Such a comparison was also made relatively easy with the Stacked Bars graph.

Finally, the Individual graphs had slightly higher performance. Both Demographic Bars and Dot Array graphs had the highest number of correct responses. Again, the Ordered Bars were not as effective, because individuals were not placed together in groups, which made the visual comparisons much more challenging. It is noteworthy that there were two correct answers for Ordered Bars, and five participants did provide one of the two correct answers. Nonetheless, even if these were counted as completely correct, the participants would have only had 47% correct answers, which is still much lower than the other *Individual* graphs.

Finally, we considered interview question 4 in which participants computed the average participation for a student in each group (i.e., Amara's participation). Only 4 of 8 visualizations could be used to answer this interview question, so we focus on participant responses to those only (see Table 6). Once again, we only see 82.4% correct responses with the Group Average Participation graph, even though the answer could directly be read from the graph (it's simply the number corresponding to any group). While this visualization does provide a lot of relevant information, participants had not seen it before and apparently many did not know how to use it.

Table 6 Performance on question 4, what was Amara's participation?

Type	Visualization	% Correct (answer given)	% Correct (strategy only)	% Correct (by type)
Derived data	Group avg participation	82.4	0	82.4
Individuals	Demographic bars	73.3	6.7	63.4
	Dot array	75	0	
	Ordered bars	35.3	0	

Table 7 Summary of accuracy (correct answer or correct strategy) of different visualization types

Question	Best answered by (based on data provided by visualization)	Accuracy (%)	Which visualization had the highest accuracy
2	Group graphs (all)	94.1	Mirror bars (100%)
3	Derived data graphs (both)	43	Dot array (68.8%)
4	Group average participation	82.4	Group average participation

Table 8 Highest levels of accuracy for different visualization types

Type	Visualization	Question(s) answered with highest accuracy	Accuracy (%)
Derived data	Equity ratio	Questions 2 and 3	23.5
Derived data	Group avg. participation	Question 4	82.4
Group	Stacked bars	Question 2	94.1
Group	Mirror bars	Question 2	100
Group	Paired bars	Question 2	88.3
Individual	Demographic bars	Question 4	80.0
Individual	Dot array	Questions 2 and 4	75.0
Individual	Ordered bars	Questions 3 and 4	35.3

In this case, of the *Individual* graphs, we see that both Demographic Bars and Dot Array had about the same level of correct answers. Again, the Ordered Bars had the worst performance as individuals were not sorted by groups.

Our analyses of participant performance showed clearly that different visualizations were better suited for different tasks. We summarize these results in Tables 7 and 8. Table 7 shows, for each question, which visualizations we would have predicted to have the highest levels of accuracy based on their design. This shows that the predicted visualization types were the most effective on questions 2 and 4, but the *Derived Data* graphs did not perform as expected for question 3, which asked about proportions of participation. This is noteworthy, because from our viewpoint, being able to make determinations about the amount of participation relative to the size of a group (i.e., a proportion) is important for understanding classroom equity and inequities. The Derived Data graphs were designed to support this thinking but were generally not as effective as hoped for.

Table 9 Which visualizations did participants find the easiest to use?

Type	Visualization	% Preferred	% Preferred (by type)
Derived data	Equity ratio	29.4	32.4
	Group avg participation	35.3	
Group	Stacked bars	47.1	27.4
	Mirror bars	17.6	
	Paired bars	17.6	
Individuals	Demographic bars	53.0	41.2
	Dot array	35.3	
	Ordered bars	35.3	

In Table 8, we provide a summary of each visualization, and for which questions it had the highest level of accuracy (including both correct answers and correct strategies). Table 8 shows that nearly all the visualizations were effective for *some* questions, except for Equity Ratio and Ordered Bars, which had low levels of accuracy across all three questions.

Visualization preferences (research question 2)

In addition to participant performance, we were interested to understand participant's subjective preferences. We provide a summary of the visualizations that participants ranked as "easiest" in Table 9. Each participant was able to nominate up to three visualizations as the easiest to interpret. Here, we provide the percentage of such nominations received by each visualization.

The most popular graph of all was the Demographic Bars graph. This is consistent with participant responses above in terms of performance, where Demographic Bars had relatively good performance across each different scenario. We also found participants preferred the Stacked Bars graph, even though its performance was more mixed. The Derived Data graphs were somewhere in the middle of preference, which aligned with their mixed results in terms of performance. Now we discuss themes for why participants preferred particular graphs.

Breaking down by individual student ($N = 12$)

Across the dataset, the most prominent feature of the graphs that participants described as helpful was the ability to see individual students. They made statements like, "individual-level information is really important" (Participant 16), because it allowed them to see "more of the data" (Participant 8) or the "raw data" (Participant 3). Especially from the perspective of answering questions about student participation, the *Individual* graphs provided the raw materials needed to make calculations in a more obvious way, whereas participants would need to work backwards from some of the other graphs to answer all the questions.

Another key reason that participants valued the Individual graphs is because they could show both the between-group variation and within-group variation. The

Individual graphs allowed you to see "what's happening with each individual student," not just at the level of the group; otherwise, you could have "one student who didn't participate at all, and then two students who participated a lot within that group" (Participant 5). What these participants described goes beyond just thinking about the graphs to answer the question prompts, but as a tool to support their thinking about classroom equity. Notably while the Ordered Bars does show both individual and group-level information, the group-level information is difficult to infer from the graph, as was evidenced by low accuracy answering questions with this visualization.

Providing the "correct response" or calculation (*N* = 10)

In contrast to the *Individual* graphs that provided the raw materials for understanding, other participants commented on how the Group Average Participation and Equity Ratio graphs already did some of the computations for them. One participant described the Group Average Participation as one of their favorites, because "was the most accessible," and it required "less calculation" (Participant 4). Another said it "would give the easiest sense of how much each group was contributing as a whole" (Participant 16).

Others said they liked the Equity Ratio graph, as it "told you the conclusion because it said whether or not it was equitable" (Participant 13). This same participant had said "equity is about things being equal for all genders and all races and that means equal access but also equal in expectation." Another participant said, "the equity ratio eliminates the error—the problem with determining if it's equitable [laughs] because it's kind of a statistic that's already been determined to measure that" (Participant 12). This participant defined equity as "enabling everyone, each and every student, to be able to maximize their potential. And so, each student might need a different opportunity or support to reach their potential."

At the same time, one participant (Participant 16) noted that she wanted to know "more about how that equity ratio was determined and, what the assumptions behind that are." We note that the participants were told how the ratio was computed, but this did not necessarily mean that they had an intuitive understanding of it. She continued it "could be an incredibly dangerous weapon" as a result of "misuse and misinterpretation, depending on what those underlying assumptions are" (Participant 16). This was an important insight, as from our perspective, the equity ratio does not represent the "correct" value, as was interpreted by some participants. It is also possible that the name "equity ratio" influenced participants' interpretations, and that a different name would have framed how participants thought differently. This same participant had defined equity as "ensuring everyone has what they need to be able to achieve what their goals are."

Side-by-side comparisons (*N* = 7)

More than any other graph, a large number of participants specifically called out the Stacked Bars as it allowed side-by-side comparisons. Participants described how they "like the Stacked Bars because they're side by side" (Participant 5), or that they could "compare the bars side by side" (Participant 3), "see the percentages side by side" (Participant 4). Notably, participants did not describe the Mirror Bars or Paired Bars in the same way, even though they were also designed to facilitate similar comparisons between percentage of participation and demographic representation.

Other ideas

Several participants ($N=4$) mentioned how the color coding helped their interpretations, but they did not articulate anything specific about a particular color scheme. One participant mentioned that she liked graphs where she could get a general gut feeling (e.g., with the Dot Array) about what was happening, without doing any calculations. Finally, one other participant mentioned that having the student contributions shown by the length of the bar (e.g., Ordered Bars, Demographic Bars) was a nice feature.

What else did people want to know?

In terms of what information was missing, or what else people wanted to know, there were far more themes, so they are broken up in terms of frequency in Table 10. The most common request ($N=6$) was for the graphs that did not have the total contributions (but just percentages) to give the total amount of participation for the class. After that, participants wanted intersectional information that was not just race ($N=4$), and information about the overall quality (i.e., different types of talk ranging from rote recall to extended explanations), not just quantity, of participation ($N=4$). Participants also mentioned that they would have found it useful to have a particular definition of equity for interpreting the graphs ($N=3$). Beyond that, a variety of ideas such as course context ($N=2$), error bars ($N=2$), or information about *which* racial groups were represented rather than generic names ($N=2$). Other one-off suggestions included the presence of accommodations (e.g., for disabled students), the total number of opportunities to participate, other forms of participation, and within-group information (for the graphs that did not have it).

Reasoning about equity (research question 3)

For the remainder of our analyses, we turn our attention away from whether participants could interpret the visualizations, to whether certain types of visualizations may have promoted different ways of thinking about equity.

Equity in general

In the beginning of the interview (Segment 1), participants were asked to define equity. Two of the participants articulated an equity as equality ideology. For example, Participant 13 said "So for me, equity is about things being equal for all genders and all races and that means equal access but also equal in expectation."

Table 10 Subcodes for missing information

Subcode	*N*	Example
Total participation	6	"In most of the graphs the total number of turns was missing, that seems like a big important piece of information"
Other demographics	4	"Other elements of identity, as opposed to just racial groups…if that was broken down and it included gender, for example"
Quality of participation	4	"I wanted to know how long the contributions were, I wanted more information about that to be able to talk about equity"
Definition of equity	2	"For some of the graphs I felt like I had more questions about what it means, means to be equitable"
Course context	2	"What's the context, is the course instructor the same? What's the prior experiences? What at what point in the term is this?"
Error bars	2	"Error bars, so that you could get an idea of the distribution"
Racial groups	2	"Some background on who group ABC and D is"
Accommodations (for disabled students)	1	"If the students had requested accommodations and if those accommodations were provided and supported the student"
Participation opportunities	1	"What is the [total] number of opportunities [to participate]?"
Forms of participation	1	"It would be wonderful to be able to complement that with a lot of other elements of participation." [e.g., nonverbal participation]
Within-group information	1	"Why in group C two people were talking a ton and then the other two people in group C were not talking at all"

Table 11 How participants categorized the visualizations as equitable or not

Type	Visualization	Yes	No	Not sure
Derived data	Equity ratio	1	12	4
	Group avg participation	4	4	9
Group	Stacked bars	5	5	7
	Mirror bars	4	5	8
	Paired bars	3	8	6
Individuals	Demographic bars	2	11	4
	Dot array	3	3	11
	Ordered bars	0	13	4
Total		22	61	53

Most participants ($N = 13$) made a general description of students getting the appropriate support or accommodations needed to succeed. Participant 4 mentioned “So equity to me means I’m providing appropriate support for each student so that may not be the same for each student. So essentially, you know, they say leveling the playing field so that may mean that some students may require more than other students.” Participant 6 stated “Equity, in terms of teaching in the classroom, means that the instructor should give access to multiple different formats, and multiple different elements of the course that allows all of the students to perform to the best of their ability.” Similarly, Participant 11 stated “I believe that equity is about ensuring that all people involved in an experience have the means and tools with which to access and benefit from that experience.”

Finally, two participants used theories of mathematics education, like Rochelle Gutierrez’s four axes of equity (Gutierrez 2002), or Complex Instruction (Cohen and Lotan 1997), to explain what they meant by equity.

Equity in specific visualizations

Moving beyond general conceptions of equity, we were interested in understanding if certain types of data visualizations prompted participants to think about equity in a particular way. Participants’ classification of participation as equitable or not is broken down by visualization in Table 11. Table 11 clearly shows that participants saw the participation associated with particular graphs as more equitable or not, compared to other visualizations. From our perspective, none of the randomly generated distributions were entirely equitable, but rather, there are aspects of participation that could be seen as both equitable and inequitable, depending on a participant’s interpretation. These forms of reasoning were what interested us most. Below, we discuss the ways in which the three different categories of visualizations prompted different types of thinking from participants.

Derived data graphs The distributions of responses for Equity Ratio and Group Average Participation were quite different. More than any other graph, participants perceived the Equity Ratio as showing what the correct answer was (all bars at one), and when the bar associated with a particular racial group did not fall on the line, participation was considered inequitable. This made the Equity Ratio graph unique from the set of visualizations. Consider Participant 12, who was steadfast in saying that for nearly all visualizations equity could not be determined without more information. However, in the case of the Equity Ratio graph she said it was inequitable, because "group B did not meet that 'one' standard." Unlike the other graphs where she felt equity was context specific, she felt as though the Equity Ratio graph told what the correct "equitable" response was. In contrast, the Group Average Participation graph did provide a computation, but no clear "right answer." As a result, people wanted to know more about the statistic to see if differences were statistically significant. Ultimately, it behaved much more similarly to the other *Group* graphs.

Group graphs When interpreting the *Group* graphs, participants often remarked that they wanted to know more about the groups. For instance, Participant 4 had a consistent reparations viewpoint, and thus, without knowing *which* racial groups were represented, equity could not be determined. In general, these graphs elicited more thinking about groups than the *Individual* graphs. When interpreting participation across the *Group* graphs, participants never brought up the idea that a single student *within* a group might be marginalized, although that type of reasoning came up often with *Individual* graphs (described below). Even if a particular group was viewed as an outlier, it was not necessarily interpreted as inequitable, because participants felt that given the right context it could potentially be equitable.

Individual graphs The *Individual* graphs evoked, not surprisingly, thinking about particular individual students. In general, participants described both Demographic Bars and Ordered Bars as *not equitable* because a single individual stood out as talking too much or not enough. As Participant 9 asked "Why did Billy never talk?" or Participant 5 noted "it seems like Casey is really domineering the conversation here." Participant 14 said participation was a "clear no" to being equitable, "because Cameron was zero." As Participant 14 elaborates,

> So unlike the other ones where I said that we needed more information to know if the distribution is equitable…we for sure know that this is not equitable and this is a new definition of or new part of the definition of equity which is to say that we are smarter together. And so because we haven't heard from Cameron at all we know that whatever kind of thing we're talking about right? And in my world it's mathematics right, so we know that we're not as smart as we would be with relate- with a relationship to the mathematics content because we haven't heard from Cameron yet. So firmly this time this is not equitable.

As these examples highlight, participants looked at the visualization, and if a single student was an outlier, they viewed it as inequitable. This is a notable difference from the *Group* graphs, wherein a single group acting as an outlier was not necessarily seen as inequitable.

Of the *Individual* graphs, the Dot Array graph evoked slightly different patterns of reasoning and had the highest number of "I don't know" responses amongst all the graphs. In general, participants did not jump to an immediate conclusion of inequity from a particular student having a very low or high level of participation. Why did the Dot Array graph not have a similar pattern? We suspect that the colors made it more ambiguous exactly how much a given participant was speaking. This contrasted the Demographic Bars and Ordered Bars, which made visually salient outliers using the length of the bars (who were either participating too much, or not enough), in a way that the Dot Array graph did not.

Conclusion and key takeaways

This study provided a variety of insights into how data visualizations can support postsecondary STEM educators to think about racial equity. This is a key first step towards developing more effective visualizations to support professional development around racial equity. When we considered how participants performed at correctly answering different questions using the visualizations, we found that what type of visualization was most effective depended on the question that was being asked. For understanding overall participation (interview question 2), the *Group* visualizations were quite effective, because this information was readily available, and the visualizations were relatively easy to interpret. For understanding the proportions of participation (interview question 3), we hypothesized the *Derived Data* visualizations would be effective because they directly provided this information, but participants did not interpret the visualizations correctly. This indicates that participants did not fully understand the computations, so there is an associated learning curve with using these, which would need to be accounted for in field studies or actual professional development. This could also be an important area for future visualization design, as helping participants understand participation relative to demographic representation is one useful way to understand participatory equity. For computing the experiences of an average student (interview question 4), most participants did realize that the Group Average Participation directly answered this question. In addition, the *Individual* graphs were the primary ones which actually provided this information. Across the board, the Demographic Bars and Dot Array Individual graphs were quite effective, as they balanced both group-level and individual-level information. Ordered Bars were much less effective, because students were not grouped together, which made it difficult to draw conclusions about particular groups of students.

There were a few key features of the visualizations that participants identified as helpful. Many participants appreciated the ability to see individual students and coordinate group-level and individual-level information. Some participants enjoyed the *Derived Data* graphs because they provided the perception of knowing the "right answer." However, with the case of the equity ratio graph we perceive this as a false confidence, which means the visualization may need to be used with caution. Also, the Stacked Bars were quite popular because people liked the physical layout of side-by-side comparisons. Participants wanted to know a lot of other relevant information, such as intersectional demographics, quality of responses, error bars, and so forth. This indicates that participants recognized the many multi-faceted areas of participation.

We also found that graphs mediated thinking about equity. At a general level, most participants described equity as something that had to be adjusted according to student needs and required providing appropriate support. However, when participants were required to make determinations about specific groups, they often reverted to an "equity as equality" ideology, because this was easier to visualize. How to avoid inadvertently promoting such equity as equality thinking is clearly a pressing challenge for the field of equitable visualization design.

Different graphs promoted different types of reasoning about equity. *Individual* graphs tended to invoke individual reasoning and were the easiest to say it was inequitable if there was an outlier. In contrast, *Group* graphs prompted group thinking, but participants wanted to know more information about the groups and the individuals within the groups. Finally, the *Derived Data* graphs simplified the process, for example, when equity ratio provided the "right answer." However, this process was not seamless, as the *Derived Data* graphs gave a false sense of correctness, and participants were not always effective at interpreting them.

There are a few key takeaways from the results above. First, the Demographic Bars graph stood out as a highly effective graph, and it was one that participants preferred. It provided individual-level information, and due to the grouping of the individual students, it was easy to make inferences about groups as well. In contrast, Ordered Bars was one of the least helpful graphs for making any inferences about groups, because individuals were not sorted together. This suggests that simply overlaying demographic information without carefully thinking about how well it can be interpreted may not be an effective strategy at all.

Second, one needs to consider how well participants can interpret the graphs. The *Derived Data* graphs had a lot of potential but had some major pitfalls. First off, these graphs created a false sense of confidence because people felt as though they knew what the answer for equity was. Equity Ratios gave the perception that a ratio of 1 was ideal, even though this would represent equality, not equity. Similarly, even in answering questions about the proportion of participation, which should have been trivial using these graphs, less than half of the responses were correct. If such *Derived Data* graphs are to be used, it is critical that participants have appropriate support.

Third, it's noteworthy that different graphs influenced how people thought about equity. Overall, participants had more nuanced ways of thinking about equity in general, but this was not easy to operationalize in terms of specific visualizations. People tended to revert to a model of "equity as equality." In general, it was much easier for participants to define inequitable rather than equitable. For example, with *Individual* graphs, if a single student was an outlier, participants typically moved to a quick judgment that participation must be inequitable. In contrast, if a particular group was an outlier, it was not necessarily seen as inequitable without more contextual information. This suggests that for any given purpose, a designer would want to use particular types of data analytics to evoke particular types of reasoning about equity from participants.

This study focused on a population of postsecondary STEM educators focused on racial equity. What remains to be seen is the extent to which these findings would extend to other contexts, such as K12 STEM education, or other types of inequity, such as gender inequity or disability inequity. Nonetheless, given that we were able to connect our quantitative findings to specific patterns of reasoning and features of the visualizations, we believe that our results should generalize across settings, at least to some extent. Exploring these further areas of inquiry with larger samples that support statistical generalization would be an important next step to this research.

Implications

This study has important implications in the form of the most effective task-visualization combinations; that is, depending on the required task, we have a better idea of which types of visualizations we should provide to achieve it. Given that different graphs mediated different ways of thinking about equity (interview question 1), it may be most effective to have a combination of an *Individual* graph (e.g., Demographic Bars) alongside a *Group* graph, which gets the focus on demographic groups. For understanding overall participation (interview question 2), any of the *Group* graphs would be quite effective. For understanding average participation by students (interview question 4), Group Average Participation and Demographic Bars were both excellent choices. The results were less clear for understanding participation relative to demographic representation (interview question 3), and this remains an area for future research.

This study also has implications for practice. First, it becomes clear that more data/information is not necessarily better. Thus, in a practical professional development context, rather than simply providing instructors with *more* data, it is important to curate *effective* data visualizations that spur the desired types of reasoning. Consider the *Derived Data* graphs for a prime example of this. Even though they could be used to directly answer interview questions, they did not necessarily provide participants with data in a form that they could easily interpret. At worst, they could lull participants into a false sense of confidence about computations that they do not really understand.

Second, a major theme was the need to coordinate individual and group-level information. This is also very important from a professional development standpoint. In real classrooms, sources of inequity arise both from within and between-group variation. Therefore, visualizations with the potential to provide information about each of these have better potential to support more nuanced thinking about in/equity. The *Individual* graphs had the most potential to do this, as they showed how particular students were participating. On the flipside, this also resulted in more individual-level thinking, and moved more away from thinking about equity in terms of groups.

Author contributions DLR was responsible for conceptualizing the study, analyzing data, and writing the first draft. SR was responsible for conducting all interviews, analyzing data, and providing feedback on the manuscript. PTS was responsible for conceptualizing the study and editing the manuscript. NS was responsible for consulting on the study and editing the manuscript.

Funding This material is based upon work supported by the National Science Foundation under Grant No. 1943146.

Data availability The data in this manuscript have not been made publicly available.

Declarations

Conflict of interest The authors have no competing financial interests related to this work.

Ethical approval This study was approved by the Institutional Review Board of San Diego State University, with approval from Rick Gulizia, Director of Research Affairs, and conducted in accordance with all relevant guidelines for participants involved.

Informed consent Informed consent was received from all participants.

References

Banes LC, Ambrose RC, Bayley R, Restani RM, Martin HA (2018) Mathematical classroom discussion as an equitable practice: effects on elementary English learners' performance. J Lang Identity Educ 17(6):416–433. https://doi.org/10.1080/15348458.2018.1500464

Beatty PC, Willis GB (2007) Research synthesis: the practice of cognitive interviewing. Public Opin Q 71(2):287–311. https://doi.org/10.1093/poq/nfm006

Cohen EG, Lotan RA (1997) Working for equity in heterogeneous classrooms: sociological theory into practice. Teachers College Press, New York

Darling-Hammond L (1998) Unequal opportunity: race and education. Brook Rev 16(2):28–32

Dimara E, Bezerianos A, Dragicevic P (2017) The Attraction effect in information visualization. IEEE Trans Visual Comput Graph 23(1):471–480. https://doi.org/10.1109/TVCG.2016.2598594

Ernest JB, Reinholz DL, Shah N (2019) Hidden competence: women's mathematical participation in public and private classroom spaces. Educ Stud Math 102(2):153–172. https://doi.org/10.1007/s10649-019-09910-w

Glaser BG, Strauss AL (1967) The discovery of grounded theory: strategies for qualitative research. Aldine Publishing Company, London

Gutierrez R (2002) Enabling the practice of mathematics teachers in context: toward a new equity research agenda. Math Think Learn 4(2 & 3):145–187

Ing M, Webb NM, Franke ML, Turrou AC, Wong J, Shin N, Fernandez CH (2015) Student participation in elementary mathematics classrooms: the missing link between teacher practices and student achievement? Educ Stud Math 90(3):341–356

Ladson-Billings G (1995) But that's just good teaching! the case for culturally relevant pedagogy. Theory Pract 34(3):159–165. https://doi.org/10.1080/00405849509543675

Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S (2012) Empirical studies in information visualization: seven scenarios. IEEE Trans Visual Comput Graph 18(9):1520–1536. https://doi.org/10.1109/TVCG.2011.279

Macefield R (2009) How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 5(1):34–45

Marbach-Ad G, McAdams KC, Benson S, Briken V, Cathcart L, Chase M, El-Sayed NM, Frauwirth K, Fredericksen B, Joseph SW, Lee V, McIver KS, Mosser D, Quimby BB, Shields P, Song W, Stein DC, Stewart R, Thompson KV, Smith AC (2010) A model for using a concept inventory as a tool for students' assessment and faculty professional development. CBE Life Sci Educ 9(4):408–416. https://doi.org/10.1187/cbe.10-05-0069

Martin DB (2019) Equity, inclusion, and antiblackness in mathematics education. Race Ethn Educ 22(4):459–478. https://doi.org/10.1080/13613324.2019.1592833

Munzner T (2014) Visualization analysis and design. CRC Press, Boca Raton

Nasir NS (2002) Identity, goals, and learning: mathematics in cultural practice. Math Think Learn 4(2–3):213–247. https://doi.org/10.1207/S15327833MTL04023_6

National Academy of Sciences (2007) Beyond bias and barriers: fulfilling the potential of women in academic science and engineering. National Academies Press, Washington

Nguyen TD, Redding C (2018) Changes in the demographics, qualifications, and turnover of American STEM teachers, 1988–2012. AERA Open 4(3):2332858418802790. https://doi.org/10.1177/2332858418802790

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

Reinholz DL, Shah N (2018) Equity analytics: a methodological approach for quantifying participation patterns in mathematics classroom discourse. J Res Math Educ 49(2):140–177
Reinholz DL, Shah N (2021) Equity and equality: how data visualizations mediate teacher sensemaking about racial and gender inequity. Contemp Issues Technol Teach Educ (CITE) 21(3):140
Reinholz DL, Wilhelm AG (2022) Race-gender D/discourses in mathematics education: (re)-producing inequitable participation patterns across a diverse, instructionally-advanced urban district. Urban Educ. https://doi.org/10.1177/00420859221107614
Reinholz DL, Stone-Johnstone A, Shah N (2020a) Walking the walk: using classroom analytics to support instructors to address implicit bias in teaching. Int J Acad Dev 25(3):259–272. https://doi.org/10.1080/1360144X.2019.1692211
Reinholz DL, Stone-Johnstone A, White I, Sianez LM, Shah N (2020b) A pandemic crash course: learning to teach equitably in synchronous online classes. CBE Life Sci Educ 19(4):ar60. https://doi.org/10.1187/cbe.20-06-0126
Reinholz DL, Johnson E, Andrews-Larson C, Stone-Johnstone A, Smith J, Mullins B, Fortune N, Keene K, Shah N (2022a) When active learning is inequitable: women's participation predicts gender inequities in mathematical performance. J Res Math Educ 53(3):204–226. https://doi.org/10.5951/jresematheduc-2020-0143
Reinholz DL, Reid A, Shah N (2022b) Not another bias workshop: using equity analytics to promote antiracist teaching. Change 54(4):11–17. https://doi.org/10.1080/00091383.2022.2078149
Saket B, Endert A, Demiralp Ç (2019) Task-based effectiveness of basic visualizations. IEEE Trans Visual Comput Graph 25(7):2505–2512. https://doi.org/10.1109/TVCG.2018.2829750
Secada WG (1989) Educational equity versus equality of education: an alternative conception. In: Secada WG (ed) Equity in education. Falmer, pp 68–88
Shah N, Lewis CM (2019) Amplifying and attenuating inequity in collaborative learning: toward an analytical framework. Cogn Instruct 1:30. https://doi.org/10.1080/07370008.2019.1631825
Shah N, Christensen JA, Ortiz NA, Nguyen A, Byun S, Stroupe D, Reinholz DL (2020) Racial hierarchy and masculine space: participatory in/equity in computational physics classrooms. Comput Sci Educ. https://doi.org/10.1080/08993408.2020.1805285
Smith MK, Jones FHM, Gilbert SL, Wieman CE (2013) The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. CBE Life Sci Educ 12(4):618–627. https://doi.org/10.1187/cbe.13-08-0154
Sukumar PT, Metoyer R (2018) A visualization approach to addressing reviewer bias in holistic college admissions. In: Ellis G (ed) Cognitive biases in visualizations. Springer, Cham, pp 161–175. https://doi.org/10.1007/978-3-319-95831-6_12
Sukumar P, Metoyer R, He S (2018) Making a pecan pie: understanding and supporting the holistic review process in admissions. Proc ACM Human–comput Interact 2(CSCW) 169(1–169):22. https://doi.org/10.1145/3274438
Sukumar P, Reinholz D, Shah N, Striegel A (2020) Visualizing participatory inequities in classroom data. OSF Preprints. https://doi.org/10.31219/osf.io/3mq6u
Tate WF (2008) "Geography of opportunity": poverty, place, and educational outcomes. Educ Res 37(7):397–411. https://doi.org/10.3102/0013189X08326409
Valdez AC, Ziefle M, Sedlmair M (2018) Priming and anchoring effects in visualization. IEEE Trans Visual Comput Graph 24(1):584–594. https://doi.org/10.1109/TVCG.2017.2744138
Yin RK (2009) Case study research: design and methods, vol 5. Sage, London
Zagallo P, McCourt J, Idsardi R, Smith MK, Urban-Lurain M, Andrews TC, Haudek K, Knight JK, Merrill J, Nehm R, Prevost LB, Lemons PP (2019) Through the eyes of faculty: using personas as a tool for learner-centered professional development. CBE Life Sci Educ 18(4):ar62. https://doi.org/10.1187/cbe.19-06-0114