Visualizing Inequities: A Step Toward Equitable Student Outcomes

Sumitra Tatapudy,[#] Rachel Potter,[#] Linnea Bostrom, Anne Colgan, Casey J. Self, Julia Smith, Shangmou Xu, and Elli J. Theobald*
Department of Biology, University of Washington, Seattle, WA 98195

ABSTRACT
The underrepresentation and underperformance of low-income, first-generation, gender minoritized, Black, Latine, and Indigenous students in Science, Technology, Engineering, and Mathematics (STEM) occurs for a variety of reasons, including, that students in these groups experience opportunity gaps in STEM classes. A critical approach to disrupting persistent inequities is implementing policies and practices that no longer systematically disadvantage students from minoritized groups. To do this, instructors must use data-informed reflection to interrogate their course outcomes. However, these data can be hard to access, process, and visualize in ways that make patterns of inequities clear. To address this need, we developed an R-Shiny application that allows authenticated users to visualize inequities in student performance. An explorable example can be found here: https://theobaldlab.shinyapps.io/visualizinginequities/. In this essay, we use publicly retrieved data as an illustrative example to detail 1) how individual instructors, groups of instructors, and institutions might use this tool for guided self-reflection and 2) how to adapt the code to accommodate data retrieved from local sources. All of the code is freely available here: https://github.com/TheobaldLab/VisualizingInequities. We hope faculty, administrators, and higher-education policymakers will make visible the opportunity gaps in college courses, with the explicit goal of creating transformative, equitable education through self-reflection, group discussion, and structured support.

INTRODUCTION

Educational Inequities in STEM

In recent decades, a growing emphasis on systemic inequities in Science, Technology, Engineering, and Mathematics (STEM) education has played a critical role in shaping current educational practices in both K-12 and higher education. Existing research on STEM education inequities has primarily focused on a set of complex structural and sociopsychological factors that contribute to the systemic educational inequities (Xie *et al.,* 2015).

Structural inequity describes a process that restricts minoritized students from accessing learning opportunities and resources. Although we primarily focus on postsecondary education in this essay, structural inequities begin earlier, particularly within K-12 STEM learning environments, where significant disparities prevail. At the K-12 level, research has identified major inequities in opportunity to take STEM courses based on student background, including their gender, race/ethnicity, and socioeconomic status (SES) (Riegle-Crumb *et al.*, 2006; Kelly, 2009; Riegle-Crumb and Grodsky, 2010). For example, only some school districts offer high-status STEM courses (e.g., college STEM preparation courses)—often conceptualized as valuable learning opportunities (Yang Hansen and Strietholt, 2018; Xu and Kelly, 2020). The concept of opportunity hoarding, where those in power maintain privilege by controlling access to opportunities, can explain the perpetuation of these systemic inequities. Past research found that advantaged families maintain social advantages

Starlette Sharp, *Monitoring Editor*
Submitted Feb 16, 2024; Revised Jul 30, 2024; Accepted Aug 27, 2024
CBE Life Sci Educ December 1, 2024 23:es9
DOI:10.1187/cbe.24-02-0086
Conflicts of interest: The authors declare no conflicts of interest.

[#]These authors contributed equally as co-first authors.
*Address correspondence to: Elli J. Theobald (ellij@uw.edu).

through actively negotiating for stratified school policies which preserve learning opportunities and resources for themselves and exclude minoritized students from advanced academic course work (Tilly, 1997; Anderson, 2010; Lewis and Diamond, 2015; Domina *et al.*, 2016). These exclusionary practices ultimately create an exclusive school system that disproportionately benefits privileged families and students (Tilly, 1997; Anderson, 2010; Lewis and Diamond, 2015; Domina *et al.*, 2016). Undoubtedly, students who consistently find themselves enrolled in advanced coursework at the K-12 level gain advantages when it comes to their postsecondary STEM experiences, including securing admission to better colleges, higher intent to declare STEM major, and greater degree attainment (Riegle-Crumb and Grodsky, 2010; Ross, 2012; Wang, 2013).

Structural inequities in K-12 education weed out minoritized students, resulting in a smaller subset attending postsecondary institutions. Strikingly, at the postsecondary level, minoritized students encounter additional and pronounced structural inequities, imposing constraints on their pursuit of STEM majors (e.g., Riegle-Crumb *et al.*, 2019; Harris *et al.*, 2020; Kim et al., 2024). These inequities not only stratify learning opportunities but also create new barriers in higher education. Recent empirical work on postsecondary STEM learning opportunities has revealed privileged groups create opportunity barriers, exclude minoritized students, and maintain their advantages in STEM attainment (Riegle-Crumb *et al.,* 2019). Particularly, despite the vast efforts of promoting equal access to learning opportunities (e.g., Algebra-for-all initiatives), minoritized students who exhibit a similar level of preparedness for postsecondary STEM are less likely to persist in and complete STEM majors, compared with their majoritized counterparts (Price, 2010; Mann and Diprete, 2013; Morgan *et al.*, 2013; Eagan *et al.*, 2014). For example, in an analysis of national representative longitudinal data of high school graduates, Morgan *et al.* (2013) found that male college students were twice as likely to enroll in a STEM major (including pre-med) than female college students, even though their average high school performance and intent toward STEM were quite similar. Students from racial and ethnic minorities who declared STEM majors also experienced lower persistence rates (Price, 2010; Eagan *et al.*, 2014), and higher rates of switching majors and dropping out (Riegle-Crumb *et al.,* 2019), compared with their White and Asian peers. Moreover, these inequities manifest in various sociopsychological aspects of STEM learning. For instance, when minoritized students witness the increased attrition of their peers and the reinforcement of cultural, gendered, and racial stereotypes, they can experience a lack of belonging or feelings of isolation, ultimately exacerbating inequities in STEM education (Leslie *et al.*, 2015; LaCosse *et al.*, 2016).

Overall, the current education system is crafted such that social inequity is maintained and exacerbated through schooling. In postsecondary education specifically, minoritized students bring their experiences of educational inequities from secondary education and confront a new, complex realm of disparities. Yet, despite what we know about the current state of educational inequities, addressing structural and systematic disparities that place certain student groups at a distinct disadvantage has been an enduring challenge in higher education. In the current study, we advance previous research on equity-minded, data-informed reflection (Bauman, 2005; Bensimon and Malcom-Piqueux, 2012; McNair *et al.*, 2020) by presenting a viable tool for data disaggregation. In addition, we contend that exhibiting visible and readable equity gaps is a small, but important, step for dismantling deficit-minded approaches and educational inequity as a whole. Here, disaggregation refers to the process by which data are processed, visualized, and analyzed not only as a whole but also by considering groupings of students by specific characteristics, such as ethnicity, SES, or any other relevant parameters. Deficit-minded approaches speak to a set of structural beliefs that ignore or even refuse to view inequitable student outcomes as consequences of gendered, racialized, and other socialized processes. Important to this paper, we emphasize that a deficit-minded approach held by instructors and administrators may pose additional obstacles to achieving equitable STEM education in higher education. Instead, we suggest an equity-minded approach that focuses on identifying and rectifying structural inequities through data-informed reflection.

Data-informed Reflection

Inequities plague our education system at all levels. For the purposes of this essay, we contend that the ultimate goal of education should be parity: equity in educational outcomes across students' multifaceted and intersecting identities including race, gender, culture, religion, first language, socioeconomic, disability, generational, and international status. We are not alone in this definition of the goal (McNair *et al.*, 2020), and we are committed to the philosophy that the ethical way to achieve parity is through the coupling of high expectations and systematic support. Note that we intend for this to be interpreted as a call for high expectations *and* systematic support. One without the other can be misconstrued as either a veneer of meritocracy or a sacrifice of rigor. Said another way, we will know our education system is equitable when there is no difference by group in course outcome (e.g., final grades, exam scores, etc.), retention, or graduation rate, and when this lack of difference is not due to disproportionate attrition. In short, education at an institution will be equitable when we have proportional representation of students (e.g., at the institution) and parity in outcomes (McNair *et al.*, 2020).

Disaggregated data are essential to understanding the extent to which there is (or is not) equity in educational outcomes. Achieving parity in educational outcomes requires working beyond identifying where inequities exist (and persist) and must include practitioners' use of disaggregated data to drive changes in practice and policy that will move classrooms, departments, and institutions closer to equity in education. McNair *et al.* (2020) present *equity-minded sensemaking* as a framework for instructors and administrators to practice critical reflection when interpreting disaggregated data. *Equity-minded sensemaking* involves analyzing educational data and policies with a focus on identifying and addressing disparities among student groups, particularly those who have been historically underserved. It emphasizes the importance of rectifying systemic inequities to create fair and inclusive learning environments for all students.

Instructors and administrators need to engage in *equity-minded sensemaking* with disaggregated data from their own classrooms and institutions in order to enact meaningful changes for current and future students. While it can be uncomfortable for instructors to engage in discussions about inequities in student outcomes, such discussions are necessary for addressing those inequities. Through reflection and collaboration with other practitioners, instructors and administrators move beyond recognizing that inequities exist and engage in a line of inquiry to better understand why the inequities persist and what actions they can take to move toward equity.

Theoretical Frameworks

We use three theoretical frameworks to understand and motivate our essay. First, we consider an Ethic of Care a prerequisite for using the data processing app we have created. Second, we hope that individuals who use the data disaggregation app will center Quantitative Critical Race Theory (or QuantCrit) when examining and interpreting the patterns they see. And third, we suggest Opportunity Hoarding as a useful framework for thinking about how to disrupt the patterns of inequities the data application reveals.

The Ethic of Care is a lens that Nel Noddings put forward that emphasizes the motivational force of caring for the success and well-being of another human, but also, when demonstrated to students, is coupled with concrete structures and supports to facilitate success (1988). An ethic of care is more than just caring, it is care that is operationalized into systems and structures in relation to students. In this case, the ethic of care refers to the care, high standards, and support instructors, departments, or institutions show toward their students (Noddings, 1988). The relational nature of an ethic of care requires instructors to place student growth and learning at the center of their work.

In many ways, one's ethic of care is the source of the motivation to pursue a job in teaching. Even for faculty for whom teaching may be secondary to their research goals, it is their ethic of care that inspires them to continually pursue pedagogical improvement. The relational nature of the ethic of care centers students and their learning as essential considerations for instructors as they work to improve their teaching practice. It is this ethic of care that we suggest the data processing app taps into. This tool can be transformative for students, vis-à-vis instructors, departments, institutions, and STEM broadly if we capitalize on the fact that instructors care and we give them data to care about. In many ways, it is this ethic of care that is a prerequisite for using the app.

Second, a QuantCrit approach recognizes that racism is structural and systemic, and rooted in educational and political systems (Stage, 2007). Essential to QuantCrit, data (in a numeric form) interpretations are neither neutral nor equitable, but can oftentimes be biased and fail to challenge underlying majoritized assumptions and power dynamics inherently embedded in quantitative approaches (Gillborn *et al.*, 2018; Strunk and Hoover, 2019). We emphasize that disparities in student academic performance should be viewed as outcomes of pedagogical practices and education policies that systemically benefit privileged groups and disadvantage minoritized student populations. In interpreting disaggregated course data, app users should always acknowledge that observed disparities among student categories, including race, ethnicity, and first-generation status, are due to an inequitable education system, rather than student preparation, aptitude, or interest.

Relatedly, to help app users conceptualize the way in which minoritized students experience an inequitable educational system, we present an opportunity hoarding framework which serves as a fundamental social mechanism that generates social inequality.[1] Essential to this framework, advantage-seeking and access-excluding collectively create an identity-based barrier between privileged students and minoritized students. In STEM education, the opportunity hoarding framework has been widely used to describe the disproportionate distribution of valuable learning opportunities across different social groups (Kelly and Price, 2011; Domina *et al.*, 2016; Riegle-Crumb *et al.*, 2019; Xu, 2023). Particularly, it elucidates how privileged groups manipulate the allocation process to maintain their advantage, thus perpetuating excessive educational inequities (Hanselman *et al.*, 2022). In this paper, we hope that app users can apply this framework and reflect on both classroom structures and social power dynamics. Specifically, understanding the way in which course design may disproportionately benefit small groups of students will enable instructors to dismantle the pattern of disparities. Additionally, it is worthwhile to note that no framework can describe the struggle of minoritized students or address educational inequity on its own. We encourage app users to explore different approaches that fit the structure of their classrooms to understand and reflect upon the educational inequities revealed by the app.

Our Essay

We have developed an R-Shiny application that facilitates the disaggregation of institutional administrative data. The interface allows authenticated users to visualize the outcomes from their course(s) and identify inequities disaggregated by student groups. The data we use as an illustrative example come from a recently published paper (Harris *et al.*, 2020). We have further anonymized the data by renaming and reordering the course numbers and adding generic course names (e.g., course 10B, 10C, etc.) and instructor names (e.g., instructor 1, instructor 2, etc.). We have edited the data (e.g., column headers, cell values of categorical variables, etc.) to be clear for users who intend to use or modify the app to maximize utility in their context. Institutions typically grant access to broadly categorized student racial demographic data (i.e., URM and non-URM; an abbreviation for Underrepresented Minority), instead of more thoroughly disaggregated data (i.e., Black, White, Latine, Asian, etc.), and often do not have categories beyond the binary (e.g., students with nonbinary gender identities are often forced to select one of two binary outcomes or "decline to respond"). However, disaggregating data to reveal inequities can motivate practitioners to pose critical questions, necessary to bridge equity gaps, and ultimately put pressure on institutions to move beyond the binary collection of identity data.

[1] In this subsection particularly, we intentionally used inequality rather than inequity to emphasize the nature of unequal distribution of Opportunities to Learn (OTL) across different student groups. Elsewhere in this paper, we used inequity to describe a wider array of disparities in STEM education.

The goal of this essay is 2-fold: First, to make freely available a modifiable R-Shiny data processing application that is designed to disaggregate student performance data by student groups. The code used to create this app is freely available in the GitHub repository here: https://github.com/TheobaldLab/VisualizingInequities. In addition, readers can explore the potential of this tool by using the demonstration app found here: https://theobaldlab.shinyapps.io/visualizinginequities/. Second, we hope to maximize the transferability of the application to user-specific contexts. Thus, the essay describes and justifies 1) key features of the app as we have constructed it; 2) how to use data and the freely available code we developed to generate these similar interfaces to visualize inequities in student success; and 3) ways in which individual instructors, groups of individuals (e.g., in departments or programs), and institutions could use such a tool to motivate more inclusive, equitable course outcomes.

Our Positionality. Positionality is important because it reifies the lenses with which we see the world and interpret our experiences (Takács, 2003). By articulating and interrogating our positionalities, we acknowledge that with these lenses and experiences come biases (Noble, 2018; Obermeyer *et al.*, 2019). We are all biased. In full, our positionalities and biases are much more than a few sentences and our identities are both socially constructed and fluid through time and space. We hope that by naming some of our current collective and individual identities, we can help the readers understand the lenses we used to approach this work.

We are an author team of individuals who are scholars, students, teachers, and researchers. We love data! Data are central to our scholarly work and often comes in the form of quantitative datasets, though we deeply appreciate and have analyzed qualitative datasets as well. Many of our methods are borrowed from the field of Ecology, where several of us trained. We extend these methods to study classrooms—which in many ways exist as ecosystems—and we strive for classes where all students are equipped with the resources and support they need to learn and thrive.

We are all currently employed by the same institution of higher education and bring to this work our collective care for students, their success, and well-being. Our work and recommendations in this essay are informed by our lived experiences in classrooms, as students and instructors, as well as by the institutional parameters of our place of employment. This is an important aspect of our positionality and how it intersects with this essay because instead of trying to be all inclusive with terms that are useful to all institutions (e.g., term, semester, trimester, quarter, etc.), we use the term that is commonplace in our context and list one other for references (e.g., department chair or head). We hope that this particular choice of terminology is not off-putting to potential users who are situated in institutions that use different language.

In addition to our collective identity, each one of us brings our unique identities and experiences to this work:

ST is a South Asian cisgender woman, geneticist, educator, and biology education researcher. She believes that cultivating self-awareness and curiosity about student perspectives is essential for creating inclusive educational spaces where all learners feel valued. To this effect, her research focuses on unveiling systemic factors that contribute to inequities and building tools that facilitate self-reflection and equity-mindedness within academic contexts.

LB is a White, nonbinary, educator who continues to benefit from existing educational policies and structures in the United States. LB previously taught science in public high schools before returning to school and completing an MEd in Educational Foundations, Leadership and Policy. In research, LB focuses on how teacher evaluation policies and practices recognize the value of teacher-student relationships on student learning and growth.

AC is a cisgendered White woman, a PhD student, and the daughter of a physics teacher. She seeks to understand how relationships shape individual experiences, equity, and resilience in both human and ecological systems.

RP is a White, Jewish, disabled, nonbinary woman with graduate-educated parents. RP currently has a master's and is pursuing a PhD in Biology. She has taught at many levels from pre-school to graduate-level courses and sees common threads for establishing a space where students feel supported, valued, inspired, and included.

CJS is a cis White woman, first-generation college student, and McNair scholar. Her life has been shaped by the belief and experience that education can be a stepping-stone from poverty to opportunity. She has benefitted from support through federal, state, and regional equal opportunity programs. She is committed to paying forward that support with her current work as an educator pursuing equality in the college classroom and as the Chair of the Faculty Council for Teaching and Learning at her institution.

JS is a cis White woman. She is a PhD student studying ecology. Her parents earned degrees beyond a college degree and she grew up with privilege. She grew up with the belief that she belonged in educational/academic settings, and she knows that this is sadly uncommon. She can recall several moments in her education that sparked her excitement and enthusiasm, and she tries to contribute to moments like that for her students as a teaching assistant.

SX is a cis Asian male social science researcher with an international background. SX used to teach highly selective high school Chemistry courses but is always a supporter of an inclusive STEM curriculum system. SX views learning as valuable opportunities yet believes its vulnerability to manipulation, where people can easily exploit opportunities in their favor, creating and exacerbating social inequities. In his research, SX is committed to understanding the structural foundations of educational inequities and breaking the opportunity barriers.

EJT is a data enthusiast, a teacher, researcher, friend, and mother. She is a cisgender, currently able-bodied White woman. As a researcher, instructor, and the current (and inaugural) chair of the Data and Analysis and Stewardship Committee for the National Society for the Advancement of Biology Education Research, she is committed to centering the student experience and understanding the human experience with data in all forms. She strives to live the goal of progress over perfection.

KEY FEATURES OF THE APPLICATION

There are several key features that are essential when creating a data processing application like the one we present here. Some features we describe as "a good idea," such as a landing page, while others are subject to federal regulations, like minimum cell size (i.e., the minimum number of observations in a disaggregated dataset). We have leveraged both best practices in the field as well as our diverse positionalities as an asset to inform key features of the application. For example, we have included user authentication to simultaneously make the information accessible and private from an instructor's standpoint. Of course, because the code for this app is freely available, all of these key features are modifiable. Below we describe the key features—in roughly chronological order, not order of importance—through the lens of advice to a user who will adapt the app to fit the needs of their own context. For ease of reading and referencing throughout the essay and figures, we have numbered the key features with roman numerals (I through VII).

I. Authentication

Authentication is the process by which verified users access the data via visualizations. Authentication is essential for securely accessing the sensitive student data used in this application. In addition, consistent with privacy regulations, authentication maintains data integrity and establishes accountability through tractable activity logs. With authenticated access, stakeholders (e.g., instructors, department chairs, deans, etc.) can individually and collectively interpret the results derived from disaggregated data, enabling a nuanced understanding of systemic challenges. This approach emphasizes proactive decision-making aimed at systemic improvements rather than punitive individual consequences.

There are two "levels" of authentication currently built into the app. First, instructors can authenticate in and view only the data from classes they teach. Because the purpose of the app is to promote self-reflection, it is most productive for the reflection to be truly directed at oneself. While some instructors may want to make comparisons between classes or contexts, these kinds of comparisons are most effective in very controlled environments. For example, a pair of instructors who both teach the introductory course may want to get together and discuss the inequities in their classes. In this case, any comparisons being made are between two trusted colleagues who both authenticate into the app to view and share their own data.

The second level of authentication is authentication for administrators. We have implemented this level of authentication in two ways: first by providing a separate script for an administrator app and second, with required authentication. When an administrator authenticates into the administrator app, they are able to view all of the classes in the dataset. This level of oversight could be helpful as the foundation of guided conversations the chair may have with individual faculty. For example, prior to a one-on-one meeting with an instructor, the chair may wish to access the visualizations for the course(s) that a particular instructor teaches. From there, the chair may probe for equity-minded reflection with the instructor. Ultimately, authentication—for both instructors and administrators—is intended to enable users (e.g., instructors, department chairs, etc.) to foster a culture of reflection and oversight that is instrumental for constructive change.

We note that authentication is not required to view the web-based demonstration of the app here: https://theobaldlab.shinyapps.io/visualizinginequities/. That is because we intend this website to be useful for illustrative purposes and to demonstrate the utility of the tool. If this tool were to be populated with local data and hosted on a website (see *Extending the Utility of the App* below), we suggest authentication in that context.

II. Landing Page

As with any website, the user is best served with a landing page to which they are directly navigated. The landing page for this data-processing app reminds users of the perspective rooted in quantitative critical theory that emphasizes the prevalence of systemic inequities that promote opportunity gaps between majoritized and minoritized students. There are no biological differences between groups of students in terms of their potential to succeed. Any differences seen in performance stem from a prevalence of systemic inequities that present hurdles to specific groups of students, making success unequally accessible.

III. Violin Plots

As with any data visualization, the creators have a choice in how to show the data. The landing page (described above) additionally informs the user how to interpret the data they are about to see (Figure 1A). While there are many ways to visualize student outcome data (e.g., boxplots, scatterplots, bar charts, etc.), the data here are visualized using violin plots. Violin plots are similar to boxplots in that they show metrics of central tendency (median and interquartile range) but are different in that they additionally display the density of the data around each value. This makes violin plots more informative than boxplots and enables users to focus not only on averages and quartiles, but also on the nuances in disparities. For example, two groups of students in a course might have a similar mean performance, but a violin plot can reveal a denser lower tail or a narrower upper tail for one group of students (Figure 1B). In more extreme cases, the distribution may be truncated, showing that students in one group do not receive the highest grades in a course, while students in another group do (Figure 1C). Considering the shape of the violin plot will move equity-minded instructors away from simply considering means, medians, and quartile ranges to build nuanced insights of inequities in course outcomes.

IV. Disaggregated Data

If data are not disaggregated, mathematically, we are only measuring the majority. Thus, disaggregating data by specific student groups is essential for understanding the learning environment (Figure 1BIV). This form of data disaggregation should move beyond simply "gap gazing," as termed by Gutiérrez (Gutiérrez, 2008), and move into trying to understand the prevailing patterns and disrupt them. Identifying

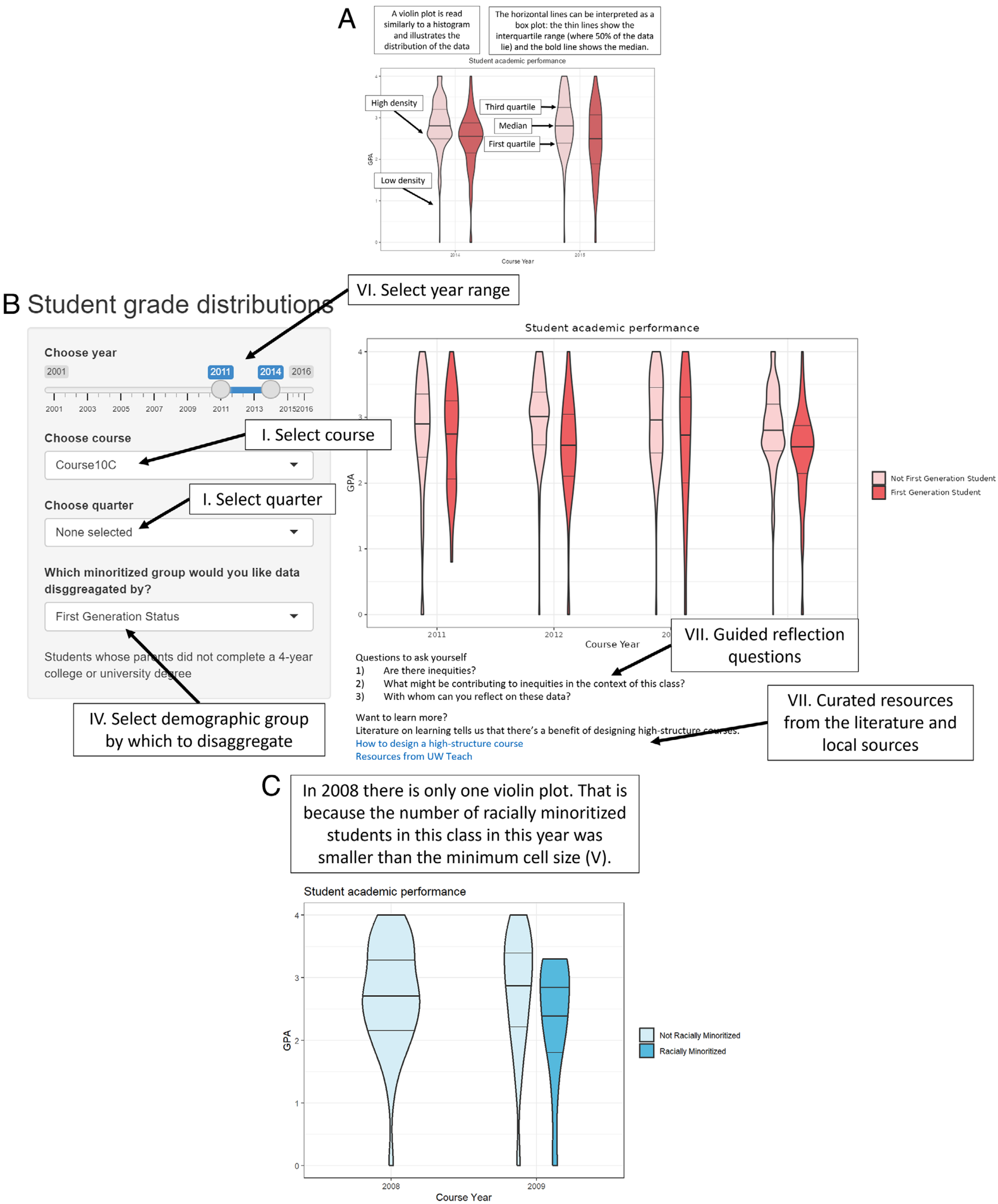

FIGURE 1. Annotated screenshots of the landing page (A), the data visualization page (B), and an illustrative example of the minimum cell size key feature (C). The annotations are numbered with roman numerals to correspond with the key features of the app as articulated in the essay.

disparities through disaggregated data analysis can catalyze data-informed self-reflection as well as targeted interventions, such as adjusting course structure (Freeman *et al.*, 2011; Haak *et al.*, 2011), centering inclusive teaching practices (Dewsbury and Brame, 2019), or institutionally aligning support and practice (Dennin *et al.*, 2017).

Choosing factors by which to disaggregate data is a critical step for users who intend to adapt the app. Specifically, some institutions may have special considerations for certain groups of students given the context. For example, approximately 30% of the students at the institution from which the illustrative data come are the first in their families to attend college. Thus, disaggregating by first-generation status is logical so that instructors can understand whether they are meeting the needs of their students. Another example, as described by McNair and colleagues (2020) is that Minnesota has a large Hmong and Somali population, so in addition to disaggregating by race as Black/African American, Latine/Hispanic, Asian, American Indian or Alaska Native, Native Hawaiian/Pacific Islander, white, and two or more races, it may be contextually relevant to determine whether students are Hmong or Somali (McNair *et al.*, 2020). As the QuantCrit framework advocates, users should move well beyond the binary perspective to understand the student experience, thus disaggregation by many identities, even simultaneously, should be a key objective (Castillo and Gillborn, 2023).

V. Establish a Minimum Cell Size

In demographics research, cell size refers to the number of observations within a cell, or in this case, the number of individuals within specific subgroups. For example, in our dataset, if there are 4 students in a class, the cell size for that class is 4. Similarly, when disaggregating data, the cell size becomes the number of observations that fit all of the criteria by which the data are being disaggregated. So, if in a single class in a single term in a single year, there are 15 students who identify as first-generation students, the cell size for this disaggregation is 15.

Disaggregation of data should be limited by a minimum cell size to maintain the data as unidentifiable. The establishment of a minimum cell size is relevant to this work because the more granularity to which the data are disaggregated, the more likely it is that an individual can be identified. For example, if there is only one woman in a given class, then cell size for women in that class is equal to 1. When disaggregating sensitive or even conceivably identifiable data, it is important to limit the cell size so that individuals cannot be identified. We call this establishing a minimum cell size.

Different organizations have different recommendations, rules, and tolerances about minimum cell size. For example, undoubtedly due to the long history of disease surveillance and public health initiatives (Fairchild *et al.*, 2007), the CDC limits minimum cell size to 100 (Center for Disease Control, 2020). There are other policies that seek to make the minimum cell size as a function not only of the numerator but also of the denominator (Wilkinson *et al.*, 2020). In education in the United States, minimum cell size is not federally regulated, aside from laws that protect the confidentiality of individually identifiable information (e.g., The Privacy Act of 1974, The Education Science Reform Act of 2002, and the US Patriot Act of 2001). The National Center for Education and Statistics (NCES) restricts minimum cell size to 3 unweighted cases by otherwise combining or resizing or recategorizing smaller cells to make it larger than 3 (NCES Statistical Standards: Standard 4-2-9 and 4-2-10, 2002). In the implementation of these federal laws in educational settings, States set their own cell size minimums. According to one source (Privacy Technical Assistance Center, 2007), "cell sizes adopted by the States range from 5 to 30 students, with a majority of States using 10 as their minimum (National Center for Education Statistics, 2010)."

For all of these reasons, (in addition to the fact that making inference on a very small number of observations is not advisable) we have chosen a minimum cell size of 10 students in the visualizations. If a user elects to visualize data that has a cell size of less than 10, the app returns a blank field—in other words, the visualization will only show the data for the group of students with more than 10 observations (Figure 1CV). We encourage users to consult with their institution's registrar or human subjects division, or the state education office when determining minimum cell size for their context. The most important thing to remember is that individual students should never be identifiable.

VI. Years Span (15 vs. 1 year)

The example dataset that we are using includes 15 years of data from one department. Longitudinal data can be useful for reflection because they illuminate trends over time. At the same time, there is an extraordinary amount of information that can be learned from one single instantiation of one single course in one single year. For example, examining data for a single year allows for a granular understanding of the variation in that 1 year. This single-year approach is particularly useful for identifying immediate challenges and considering targeted interventions. On the other hand, viewing data over a 15-year period provides a comprehensive, long-term, and historical perspective. This extended time frame allows for the identification of persistent patterns, trends, and systemic issues that may not be immediately evident in shorter-term analyses.

In the app, we allow users to select any year range that they prefer (Figure 1BVI). Undoubtedly some users will focus only on a year or 2 while others will want to take a longer view, and yet others will want to look both at the long view and the short term (by toggling between views). Allowing this flexibility has greater potential for app utility.

VII. Intentional Pairing with Resources

This is a tool for self-reflection but we want to emphasize the intentional pairing of guided reflection and institutional support. In the app, we have chosen to link only a handful of resources (instead of a large list) to prevent the feeling of utter overwhelm (Figure 1BVI). In addition, we have intentionally linked a few sources to the literature on evidence-based teaching practices but also, importantly, linked a few sources that are internal to our institution. For example, Centers for Teaching and Learning can be incredible resources to help instructors reflect on and improve upon their teaching practice. Finally, promoting group reflection and group problem solving can enhance collaborations across campus.

UTILITY OF THE APP AND THREE SCENARIOS

We anticipate this app will be useful to many people in institutions of higher education. First, we hope that seeing these data disaggregated in this way prompts instructors to use data to self-reflect on their courses and their students' experiences. Next, we hope that groups of instructors, whether in a course series, a department, or a program will find this tool useful for group reflection and professional support. And finally, we envision institutions using a tool like this to more intentionally support instructors in their teaching practice. To explore the utility of this app across these different users (which we will refer to as instructors, departments, and institutions), we first present three reflection questions and then present three scenarios in which this reflection could be productive.

Reflection Questions

We suggest three questions for consideration as the data in the app are being explored: 1) Are there inequities in student grades? 2) What might be contributing to these inequities in the context of this or these classes? 3) With whom can you reflect on these data? Below we elaborate on these three questions to operationalize them into equity-minded, reflective teaching practices.

Are there Inequities in Student Grades? When engaging with disaggregated student outcome data (such as the data presented in the app), users might be tempted to rely on quantitative metrics like significance tests to answer the question "are there inequities." However, significance tests for the data presented in the app have several limitations. For example, because students from minoritized groups are underrepresented, small sample sizes (e.g., especially from small classrooms) may result in statistically insignificant inequities when meaningful inequities are experienced. Conversely, tiny inequities in large enrollment classes may have infinitesimally small *p*-values simply because the sample size is so large. Thus, we opted not to generate or provide the results from significance tests in the app. This is a concrete way to move beyond relying on *p*-values and single metrics of magnitude when identifying inequities in student outcomes (see also Gillborn *et al.*, 2018 for further argument about the problematic use of *p*-values in equity research). Said another way, all manifestations of inequity deserve reflection and action, regardless of statistical significance. Therefore, equity-minded practitioners must identify where there are inequities in student grades with the intention of prioritizing courses and programs with the highest need but ultimately addressing all disparities.

To effectively prioritize, we recommend that users concentrate on identifying the context with the greatest inequity. To do this, users can explore the data, and make several comparisons. For example: 1) between students with different identities within a single class, 2) in a single class across time (e.g., years, quarters), and 3) between different classes.

1. Comparisons within a single class: Figure 2A demonstrates how users can explore inequities within a single class between students from different demographic groups, disaggregating by binary gender (panel 1), racial minoritization (panel 2), and parental education (panel 3). One reasonable prioritization would be to understand while all minoritized groups experienced inequities, racially minoritized and first-generation students experienced larger inequities (seen by comparing the medians). Furthermore, because violin plots move beyond medians, it becomes apparent that the density in the highest grade band is narrower for racially minoritized students compared with racially majoritized students, demonstrating that racially minoritized students are underrepresented at the highest achievement levels.
2. Comparisons within a single class across time: When instructors identify a particular demographic group facing large inequities, they may also choose to examine that demographic group in a single course across different academic years. As shown in Figure 2B, by looking at year-to-year differences in inequities (e.g., from 2004 to 2007), instructors may be able to determine whether the trend in outcome disparities is moving toward equity, inequity, or remaining unchanged. In this case, the data suggests a persistent inequity of similar magnitude each year. By analyzing trends across multiple years, instructors can gather more robust evidence of systemic inequities, as single-year data might be skewed by anomalies or obscured by noise.
3. Comparisons between classes: Figure 2C shows course-to-course differences in outcome disparities by visualizing three different courses taught by the same instructor (e.g. Course 1 in panel 1, Course 2 in panel 2, and Course 3 in panel 3). With this comparison, the instructor can prioritize reform in the course with the largest inequity. For example, in Figure 2C panel 2: Course 2 exhibits larger disparities compared with Course 1. In Course 3, the sample size of minoritized students was fewer than 10, thus only majoritized students are displayed in the analysis. The absence of data from a student group emphasizes the potential gravity of inequities, highlighting the need for instructors, groups of instructors, and/or departments to collectively reflect on the systems and structures that may be preventing students from enrolling in these classes. Ultimately, comparing inequities between courses can help an instructor prioritize their efforts in the short term.

It is essential for equity-minded users to view app data not as the sole indicator of equitable outcomes but as a starting point for self-reflection and educational reform. Addressing all inequities is crucial and achieving equitable outcomes requires persistent reflection, effort, and reform across institutional levels. Through exploratory and iterative comparisons as described above, instructors, departments, and institutions can identify inequities to prioritize addressing, and thereby engage in the pursuit of progress.

What Contributes to Inequities in this or these Classes? After identifying inequities, instructors, departments, and institutions should consider contextual factors associated with inequities in student outcomes. The data provided in the app facilitate comparisons in inequities in student outcomes within the same class, across different classes taught by the same instructor, and over time (e.g., quarters or years). Once these inequities are identified, the next step is to reflect on what contributes to these inequities. Are there course fea-

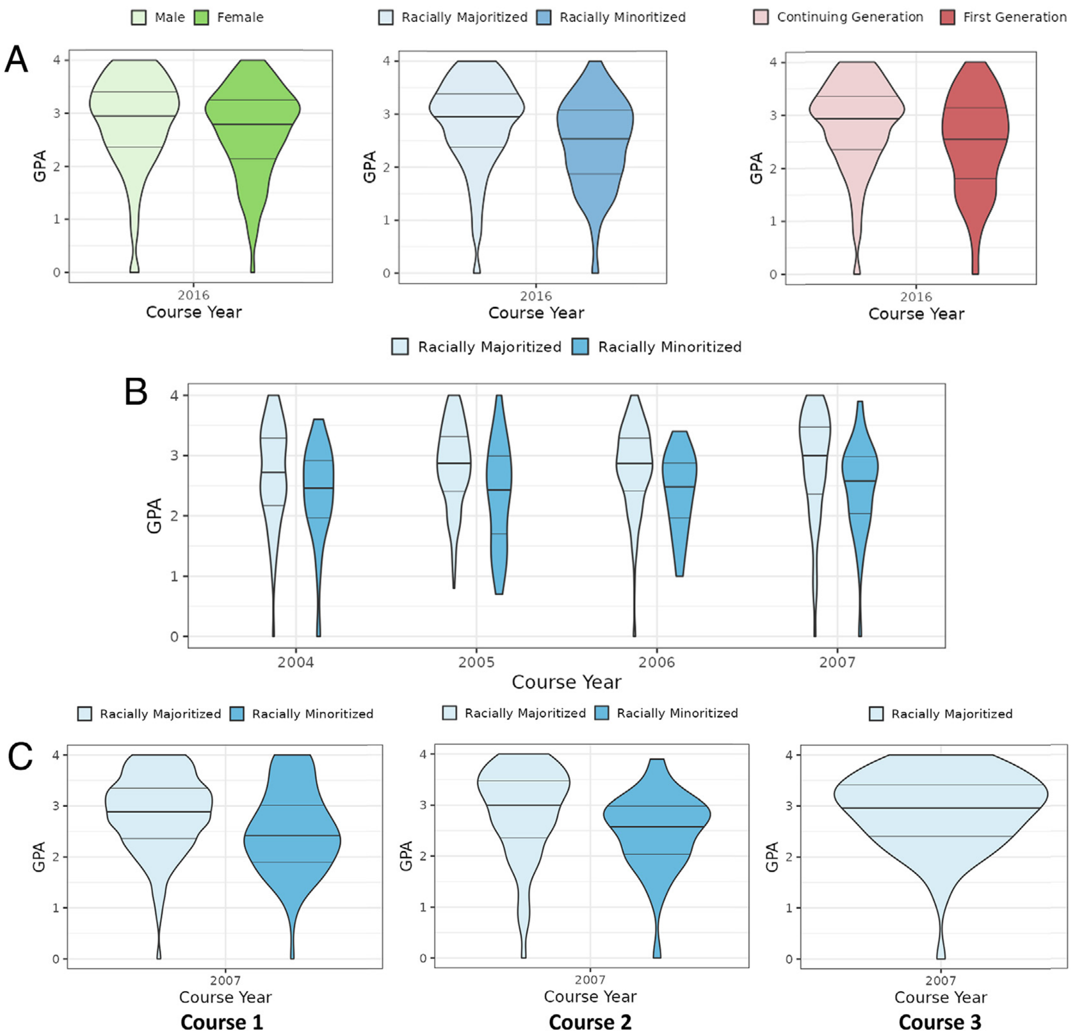

FIGURE 2. To answer the question "Are there inequities in student grades?" instructors should consider exploring trends in student outcomes by course, year, and student group. (A) The largest inequities in this class exist for students from racially minoritized groups and first-generation backgrounds as opposed to inequities by binary gender. (B) The inequity between racially majoritized and racially minoritized students persists from 2004 to 2007. (C) There are larger inequities in student outcomes in Course 2 than in Course 1. For course 3, the sample size of minoritized students was < 10 so only majoritized students are displayed.

tures that impede equity? Are there interventions or course modifications that could improve equity? For example, active learning classes (Theobald *et al.*, 2020) and high structure classes (Haak *et al.*, 2011) are both disproportionately beneficial for racially minoritized students as well as students who are the first in their families to attend college. Similarly, values affirmation exercises (Miyake *et al.*, 2010) as well as interactive course designs (Lorenzo *et al.*, 2006) have been shown to reduce the performance inequities between men and women in physics classes. Additional contextual factors to consider include disproportionate attrition through introductory sequences prior to upper division courses (e.g., Harris *et al.*, 2020) and specifically considering whether prerequisite courses are disproportionally filtering out minoritized students (e.g., Kiser *et al.*, 2022). Reflecting on the contextual factors that contribute to inequity can provide valuable insights for instructors as they consider curriculum (re)design, for departments reflecting on strategies to combat minoritized student attrition, and for institutions allocating resources to departments to facilitate equity-minded initiatives.

With Whom can you Reflect on these Data? Instructors are not alone in working toward equity-minded reflection and instruction. Are there other instructors with whom to reflect on these data? Are there experts at local centers for teaching and learning with whom to reflect on these data? Are there colleagues nationally who teach similar courses with whom to reflect on these data?

The choice of with whom to reflect can build momentum for a culture of change toward building more equitable classrooms. For instance, as an administrator, reflecting with other administrators might yield important decisions in different ways than reflecting with instructors. Similarly, reflecting on the data with other instructors, student advisors and/or administrators should focus on building systems to enable changes for equity. This collaborative approach fosters a collective pursuit of equity-mindedness and encourages a supportive environment for transformative change.

Understanding inequities is not simple; inequities stem from many interconnected factors spanning classrooms, departments, institutions, and society at large. The presence of inequities serves as a reflection of systemic issues rather than individual shortcomings. Therefore, while individual instructors possess agency and can drive change within their classrooms, there is an equal responsibility on departments and institutions to foster a culture of change. To further operationalize these ideas, we present three examples that attune to individual instructors, groups of instructors (departments), and institutions.

Three Scenarios

Scenario 1: Individual Instructors. Instructors can use this app to reflect on student outcomes and student experiences in their courses. Whether instructors engage in self-reflection on their own or with others, they should activate their ethic of care, and engage with this work using a quantitative critical race theory perspective. Furthermore, instructors should imbue a growth mindset for students and for themselves (see narrative on landing page). By critically reflecting on the inequities present, instructors will build their *equity-minded sensemaking* (McNair *et al.*, 2020) and with time can work toward solutions.

In addition to engaging in this self-reflection to understand inequities in student outcomes, instructors should use the data to make evidence-based instructional decisions to eliminate inequities and support the students they are currently teaching. Instructors can explore the data from their own classes and practice interpreting data with a focus on the structural factors that perpetuate inequitable course outcomes. In addition, even though the data come from a single instructor's class, students, classes, and instructors are embedded within a broader context: departments, programs, institutions, and disciplines. All of these contexts influence course-level outcomes so instructors should attune to these layered factors during their self-reflection. While we recognize that an individual instructor cannot address systemic problems at all levels, small scale interventions starting in the classroom can have meaningful impacts and may inspire larger scale changes.

Ultimately, this app provides instructors with a starting point to develop their equity-minded, data-informed reflection and curates resources to guide decisions about instructional practices and policies.

Scenario 2: Groups of Instructors (Teaching Teams or Departments). Groups of instructors can engage with this app in multiple ways, including informal conversations around the patterns they see in the data to more structured work to align the student experience in a department. Informal conversations with colleagues can help individuals become more comfortable talking about student outcome data, systemic disparities, and inequities. Eventually this work can move to more formal settings, like Professional Learning Communities (PLCs). When engaging in difficult conversations, it is important that groups of instructors set norms and expectations for how to participate in group reflection. We suggest practices such as viewing reflection as a learning opportunity, discussing ideas and not people, and being nonjudgmental and respectful of effort, among others.

An example of how a PLC could use this app to engage in conversations would be for teaching teams to look at student outcome data across all sections of a single course. Alternatively, a PLC that includes instructors who teach similar content and skills but across divisions (e.g., introductory and upper division) might find strength in reflecting on these data together. Another, more tailored approach which could be particularly beneficial in circumstances where facilitating open discussions with raw data may pose challenges, could involve the department chair or administrator facilitating one-on-one conversations with instructors or teams of instructors from classes with substantial equity gaps. It would be ideal if this department chair or administrator were trained in facilitating difficult conversations and had experience working intentionally to overcome equity gaps in their courses. This is an intentional support an institution could provide to department chairs. Alternatively, a third-party expert, for example a facilitator from a center for teaching and learning, a facilitator from an organization specializing in facilitating conversations around equity, and/or a DBER researcher with expertise in course design for equity could all be partners in facilitating this conversation between a department chair and an instructor. We urge leaders to seek this training or expertise both within and beyond the department. Ultimately, these kinds of conversations could serve as a basis for targeted interventions and support for instructors, such as mentoring, structured feedback, coteaching, or coaching. This "small group first" approach allows for a focused and constructive conversation around addressing disparities and supporting individual instructors.

Departmental leaders who want to formalize the use of the app should ask themselves: what departmental policies and practices would facilitate data-informed group reflection? For instance, departmental leaders could implement a policy of scheduling regular data review sessions during faculty meetings or departmental retreats (with the hope that individuals would break out into smaller groups for initial conversations). Alternatively, data from this app may be useful as the foundation of peer conversations and potentially collegial evaluations. It is important to remember that the end goal is parity in student outcomes, but it is the process of reflection paired with intentional action that will get us there. Thus, it is this reflection and this action that should be the substance of the conversation with peer mentors. Finally, departmental and institutional leaders can value this reflection by requiring evidence of reflection, iteration, and progress in merit and promotion packages.

Scenario 3: Institutions. Evaluating teaching at an institutional level is often fraught with fear and anxiety on the part of faculty. At the same time, the current tools for evaluating teaching, on which we rely heavily (e.g., Student Evaluations of Teaching) are biased, systematically disadvantage instructors from minoritized groups, and are not correlated with student learning (Kreitzer and Sweet-Cushman, 2022). To address this, many institutions are moving away from a classic form of teaching evaluation that relies solely on student and peer evaluations of teaching, and instead turning to a more holistic approach to evaluation. A holistic approach to evaluation includes opportunities for faculty to discuss professional growth and identify areas of their teaching for improvement. To achieve this level of professional development, faculty must be participating in regular reflection on their teaching practice and their efficacy in the classroom. Many current models of evaluations of teaching do not support a true cycle of formative assessment of teaching, instead relying solely on annual summative external assessments of faculty performance.

There are many institutional barriers in the transition to a holistic evaluation approach. One barrier is easy access to digestible data for faculty to use in their self-evaluations and in promotion and tenure packages. At most institutions, the burden of curiosity and data processing lies solely on the faculty member. This app helps address this need. In addition, it is important for institutions to set goals, metrics of success, and best practices, then provide faculty evidence of their performance. The activation energy for individual faculty to use data-driven reflection in their self-assessment of teaching is too high and is too variable across disciplines. Access to disaggregated data should be a service provided by the institution.

An additional barrier is trust that identification of areas of improvement will not immediately result in punitive action. A culture of support and growth of teaching, rather than punishment, is critical for adoption of new ideas and practice by faculty. In practice, it is important that faculty voices contribute to the development of alternative review processes in collaboration with academic administrators. Policies developed by faculty governance, and/or faculty unions are practical systems to provide checks and balances for the use of these types of data in evaluation efforts.

At an institutional level, availability of these types of data could provide a roadmap for professional development programs and services at the departmental and college level to support faculty in course redesign or classroom practice changes. When faculty identify areas for improvement, institutions need to be prepared to offer active support networks, training, and compensation to faculty engaged in active improvement of their courses. It is important to note, however, that interventions often require iteration and refinement (e.g., Casper *et al.*, 2019), and calibration on expected effect sizes is often critical (Kraft, 2020). Finally, the evaluation of teaching must also recognize the persistent nature of inequities in the system and value the ongoing efforts to mitigate inequities. A pitfall of widespread access to these data could be the temptation to use them solely to deny tenure, promotion, or merit to faculty. Ultimately, such punitive action undermines the trust and collaboration needed for continuous improvement of teaching.

EXTENDING THE UTILITY OF THE APP

This data processing application is just one step for progress toward equity in higher education. We need a radical culture shift wherein equity is prioritized. To that end, there are several next steps and many additional considerations for individuals, groups of individuals, and institutions in their institutionalization of this culture shift. Here we articulate some additional considerations as they intersect with the app we present here.

Future Modifications to the App

First, we acknowledge the limitations of visualizing inequities between binary demographic groups. Identities exist on continuums and multiple continuums simultaneously. Reducing variation to binaries not only dehumanizes students' experiences but also does not accurately reflect the variation in classrooms and classroom experiences. Thus, we urge individuals and institutions to collect data encompassing a wide range of identities beyond binary distinctions for racially minoritized students, gender, and other demographic factors. For example, data collection must move beyond the traditional binary of "male" and "female," to include groupings such as nonbinary and genderqueer identities, and analyses and disaggregation should incorporate multiple and intersectional identities. It is not until we recognize the diverse students in our classrooms and their diverse experiences that we can truly achieve equity.

Host the Tool Centrally

In a department or institution where obsessive pursuit of equity is the norm and disaggregation of data is common practice, a tool like this one would be freely available to all instructors. In that case, the institution or department may consider working with IT to host the application in a centralized location. To do this, we envision a collaboration between IT (either central IT or departmental IT) and departmental leadership to transition the app out of R-Shiny, creating a more user-friendly, web-integrated interface accessible to instructors and administrators without coding expertise. From here, there may be opportunities to sync the app with institutional data repositories, course dashboards for instructors, or to learning management systems to provide additional reporting for instructors.

With this practice, Key Feature I (authentication) will become increasingly important. The goal of this app is not to view or judge others' instruction; the goal is self-reflection about how to better serve the students in one's class.

Pair Course Outcomes with Additional Data

The course-level outcome data around which the application was built is useful in itself, but there may be additional sources of data that a user may consider collecting and either integrating into the app or using to supplement the guided reflection. For example, we know that measures of prior academic performance, such as high school Grade Point Average (GPA) or SAT scores, or course grades in previous college courses, are correlated with current outcomes (Wang, 2013). Because of this, it might be useful to collect these data and integrate them with the data in this app so that an instructor can look at course outcomes controlling for and not controlling for prior performance. If prior performance data are integrated, we encourage users to continue interrogating "raw" course outcomes (i.e., not controlling for prior performance) because it is a reflection of what students experience in the class. Said another way, students are not fitting statistical models in their heads to control for high school GPA and grade in the previous course. Instead, they are sitting in class experiencing inequities in raw outcomes.

In addition to data on academic performance, there are other sources of data that can help inform equity-minded, data-informed reflection. For example, it can often be fruitful to pair data on student outcomes such as grades, exam scores, and persistence with data about instructional practices such as noncontent related talk (Seidel *et al.*, 2015), active learning strategies (Smith *et al.*, 2013; Eddy *et al.*, 2015), volume in the class (Owens *et al.*, 2017), or even self-reflection on such features (Wieman and Gilbert, 2014). There are numerous studies demonstrating the utility of active learning or increased course structure on student outcomes (Eddy and Hogan, 2014; Freeman *et al.*, 2014; Moon *et al.*, 2021) and equity (Eddy and Hogan, 2014; Haak *et al.*, 2011; Theobald *et al.*, 2020), so pairing student outcomes with classroom practices is a logical next step.

Finally, there is mounting evidence that the self-reported affect measures of students' classroom experiences impact learning outcomes. For example, Eddy and Hogan (2014) report that in classes with higher structure, students have a greater sense of community with each other. Additionally, Ballen (2020) report that in the transition from passive learning to active learning, students' sense of self-efficacy and sense of belonging increased. These studies demonstrate that the instructor can impact affect and outcomes by increasing the amount or type of active learning, so pairing data on self-reported affect with student outcome data could be an informative practice.

Ultimately, there are many sources and types of data that might be useful to pair with the data shown in the application. For now, the app is designed to reveal inequities in student course outcomes, but this app is not intended to limit the types and sources of data instructors use for self-reflection.

CONCLUSIONS

The application presented here disaggregates course-level student outcome data. Paired with curated resources for intervention, we urge instructors to adopt an equity-minded, data-informed reflective teaching practice (McNair *et al.*, 2020). Using frameworks, including ethics of care (Noddings, 1988), QuantCrit (Gillborn *et al.*, 2018), and opportunity hoarding (Tilly, 1997), instructors can understand the inequities in their classrooms with the intention of disrupting them. In other words, instructors have the capacity to address institutional issues within their classrooms, but they should not stop there. Ultimately, and together, we can reify equity in postsecondary STEM classrooms.

ACKNOWLEDGMENTS

We would like to thank Sara Berk, Jake Cooper, Jon Herron, Aji John, Erika Offerdahl, and two anonymous reviewers for their thoughtful comments on the manuscript and the data visualization app. Sumitra Tatapudy is a Washington Research Foundation postdoctoral scholar, and we are grateful for the foundation's support in pursuing this work.

REFERENCES

Anderson, E. (2010). The imperative of integration. Retrieved January, 2024 from https://ebookcentral.proquest.com/lib/washington/reader.action?docID=557135

Aranda, M. L., Diaz, M., Mena, L. G., Ortiz, J. I., Rivera-Nolan, C., Sanchez, D. C., ... & Tanner, K. D. (2021). Student-authored scientist spotlights: Investigating the impacts of engaging undergraduates as developers of inclusive curriculum through a service-learning course. *CBE Life Sciences Education*, *20*(4), ar55.

Ballen, C. J. (2020). Enhancing diversity in college science with active learning. *Active Learning in College Science: The Case for Evidence-Based Practice*, 873–887.

Bauman, G. L. (2005). Promoting organizational learning inhigher education to achieve equity in educational outcomes. *New Directions for Higher Education*, *2005*(131), 23–35.

Bensimon, E. M., & Malcom-Piqueux, L. E. (2012). *Confronting Equity Issues on Campus: Implementing the Equity Scorecard in Theory and Practice*. New York, NY, USA: Routledge, Taylor & Francis Group.

Casper, A. M., Eddy, S. L., & Freeman, S. (2019). True Grit: Passion and persistence make an innovative course design work. *PLoS Biology*, *17*(7), e3000359.

Casper, A. M. A., Rebolledo, N., Lane, A. K., Jude, L., & Eddy, S. L. (2022). "It's completely erasure": A qualitative exploration of experiences of transgender, nonbinary, gender nonconforming, and questioning students in biology courses. *CBE—Life Sciences Education*, *21*(4), ar69.

Castillo, W., & Gillborn, D. (2023). *How to "QuantCrit:" Practices and questions for education data researchers and users*. (EdWorkingPaper: 22-546). Retrieved January, 2024 from Annenberg Institute at Brown University.

Center for Disease Control. (2020). 2019 YRBS National, State, and District Combined Datasets User's Guide. Retrieved January, 2024 from https://www.cdc.gov/healthyyouth/data/yrbs/pdf/2019/2019_YRBS_SADC_Documentation.pdf

Daane, A. R., Decker, S. R., & Sawtelle, V. (2017). Teaching about racial equity in introductory physics courses. *Physics Teacher*, *55*(6), 328–333.

Dennin, M., Schultz, Z. D., Feig, A., Finkelstein, N., Greenhoot, A. F., Hildreth, M., ... & Miller, E. R. (2017). Aligning practice to policies: Changing the culture to recognize and reward teaching at research universities. *CBE Life Sciences Education*, *16*(4), es5. https://doi.org/10.1187/cbe.17-02-0032

Dewsbury, B., & Brame, C. J. (2019). Inclusive teaching. *CBE Life Sciences Education*, *18*(2), fe2.

Domina, T., Hanselman, P., Hwang, N., & McEachin, A. (2016). Detracking and tracking up: Mathematics course placements in California Middle Schools, 2003–2013. *American Educational Research Journal*, *53*(4), 1229–1266.

Eagan, K., Hurtado, S., Figueroa, T., & Hughes, B. E. (2014). Examining STEM pathways among students who begin college at four-year institutions. Retrieved January, 2024 from https://scholarworks.montana.edu/xmlui/bitstream/handle/1/15115/Hughes_NAS_white_2014.pdf?sequence=1

Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE Life Sciences Education*, *13*(3), 478–492.

Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large Science, Technology, Engineering, and Mathematics classes. *CBE Life Sciences Education*, *14*(2), 14:aR23.

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE Life Sciences Education*, *13*(3), 453–468.

Fairchild, A., Bayer, R., & Colgrove, J. (2007). Privacy and public health surveillance: The enduring tension. *The Virtual Mentor: VM*, *9*(12), 838–841.

Teaching@UW; University of Washington. First-generation college students. Retrieved October 28, 2019 from https://teaching.washington.edu/inclusive-teaching/supporting-specific-student-groups/first-generation-students/

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

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE Life Sciences Education*, *10*(2), 175–186.

Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, "Big Data" and principles for a critical race theory of statistics. *Race Ethnicity and Education*, *21*(2), 158–179.

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

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213–1216.

Hanselman, P., Domina, T., & Hwang, N. (2022). Educational inequality regimes amid Algebra-for-All: The provision and allocation of expanding educational opportunities. *Social Forces*, *100*(4), 1722–1751.

Harris, R. B., Mack, M. R., Bryant, J., Theobald, E. J., & Freeman, S. (2020). Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a "hyperpersistent zone". *Science Advances*, *6*(24), eaaz5687.

Teaching@UW; University of Washington. Inclusive teaching. Retrieved March 6, 2023 from https://teaching.washington.edu/inclusive-teaching/

Kelly, S. (2009). The Black-White gap in mathematics course taking. *Sociology of Education*, *82*(1), 47–69.

Kelly, S., & Price, H. (2011). The correlates of tracking policy: Opportunity hoarding, status competition, or a technical-functional explanation?. *American Educational Research Journal*, *48*(3), 560–585.

Kim, J., Soler, M., Zhao, Z., & Swirsky, E. (2024). *Race and Ethnicity in Higher Education: 2024 Status Report*. Washington, DC: American Council on Education.

Kiser, S. L., Andrews, C. M., Seidel, S. B., Fisher, M. R., Wright, N. A., & Theobald, E. J. (2022). Increased Pass Rates in Introductory Biology: Benefits and Potential Costs of Implementing a Mathematics Prerequisite in a Community College Setting. *CBE—Life Sciences Education*, *21*(4), ar72.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253.

Kreitzer, R. J., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, *20*(1), 73–84.

LaCosse, J., Sekaquaptewa, D., & Bennett, J. (2016). STEM stereotypic attribution bias among women in an unwelcoming science setting. *Psychology of Women Quarterly*, *40*(3), 378–397.

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

Lewis, A. E., & Diamond, J. B. (2015). *Despite the Best Intentions: How Racial Inequality Thrives in Good Schools*. New York, NY, USA: Oxford University Press.

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics*, *74*(2), 118–122.

Mann, A., & Diprete, T. A. (2013). Trends in gender segregation in the choice of science and engineering majors. *Social Science Research*, *42*(6), 1519–1541.

McNair, T. B., Bensimon, E. M., & Malcom-Piqueux, L. (2020). *From Equity Talk to Equity Walk: Expanding Practitioner Knowledge for Racial Justice in Higher Education*. Hoboken, NJ, USA: John Wiley & Sons.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, *330*(6008), 1234–1237.

Moon, S., Jackson, M. A., Doherty, J. H., & Wenderoth, M. P. (2021). Evidence-based teaching practices correlate with increased exam performance in biology. *PLoS One*, *16*(11), e0260789.

Morgan, S. L., Gelbgiser, D., & Weeden, K. A. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research*, *42*(4), 989–1005.

Morton, T. R., Agee, W., Ashad-Bishop, K. C., Banks, L. D., Barnett, Z. C., Bramlett, I. D., ... & Woodson, A. N. (2023). Re-envisioning the culture of undergraduate biology education to foster Black student success: A clarion call. *CBE Life Sciences Education*, *22*(4), es5.

National Center for Education Statistics. (2010). Statistical methods for protecting personally identifiable information in aggregate reporting. *2011*(603). Retrieved January, 2024 from https://nces.ed.gov/pubs2011/2011603.pdf

NCES Statistical Standards: Standard4-2-9 and 4-2-10. (2002). Retrieved January, 2024 from https://nces.ed.gov/statprog/2002/std4_2.asp

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. Retrieved January, 2024 from Https://psycnet.apa.org › Record › 2018-08016-000https://psycnet.apa.org › Record › 2018-08016-000, 229 https://psycnet.apa.org/fulltext/2018-08016-000.pdf

Noddings, N. (1988). An ethic of caring and its implications for instructional arrangements. *American Journal of Education*, *96*(2), 215–230.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453.

Owens, M. T., Seidel, S. B., Wong, M., Bejines, T. E., Lietz, S., Perez, J. R., ... & Tanner, K. D. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(12), 3085–3090.

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*(6), 901–910.

Privacy Technical Assistance Center. (2007). Dispute avoidance and resolution service (DARS), frequently asked questions. Retrieved January, 2024 from https://studentprivacy.ed.gov/sites/default/files/resource_document/file/FAQs_disclosure_avoidance_0.pdf

Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The role of gender and friendship in advanced course taking. *Sociology of Education*, *79*(3), 206–228.

Riegle-Crumb, C., & Grodsky, E. (2010). Racial-ethnic differences at the intersection of math course-taking and achievement. *Sociology of Education*, *83*(3), 248–270.

Riegle-Crumb, C., King, B., & Irizarry, Y. (2019). Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields. *Educational Researcher*, *48*(3), 133–144.

Ross, T. (2012). Higher education: Gaps in access and persistence study. Retrieved January, 2024 from https://nces.ed.gov/pubs2012/2012046.pdf

Seidel, S. B., Reggi, A. L., Schinske, J. N., Burrus, L. W., & Tanner, K. D. (2015). Beyond the biology: A systematic investigation of noncontent instructor talk in an introductory biology course. *CBE Life Sciences Education*, *14*(4), ar43.

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE Life Sciences Education*, *12*(4), 618–627.

Stage, F. K. (2007). Answering critical questions using quantitative data. *New Directions for Institutional Research*, *2007*(133), 5–16.

Strunk, K. K., & Hoover, P. D. (2019). Quantitative methods for social justice and equity: Theoretical and practical considerations. In K. K. Strunkand & L. A. Locke (Eds.), *Research Methods for Social Justice and Equity in Education* (pp. 191–201). Cham, Switzerland: Springer International Publishing.

Takács, D. (2003). How does your positionality bias your epistemology. *Thought and Action*, Retrieved January, 2024 from https://www.semanticscholar.org/paper/2cd1c486422c0d49a1eb4b0876e9a03a99da8d2d

Teaching @ UW; University of Washington. Teaching@UW. Retrieved June 29, 2023 from https://teaching.washington.edu/

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(12), 6476–6483.

Tilly, C. (1997). Durable inequality. Retrieved January, 2024 from https://people.duke.edu/~jmoody77/FacFav/tillyChap.pdf

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, *50*(5), 1081–1121.

Wieman, C., & Gilbert, S. (2014). The teaching practices inventory: A new tool for characterizing college and university teaching in mathematics and science. *CBE Life Sciences Education*, *13*(3), 552–569.

Wilkinson, K., Green, C., Nowicki, D., & Von Schindler, C. (2020). Less than five is less than ideal: Replacing the "less than 5 cell size" rule with a risk-based data disclosure protocol in a public health setting. *Canadian Journal of Public Health. Revue Canadienne de Sante Publique*, *111*(5), 761–765.

Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual Review of Sociology*, *41*, 331–357.

Xu, S., & Kelly, S. (2020). Re-examining the Public–Catholic School gap in STEM opportunity to learn: New evidence from HSLS. *Social Sciences*, *9*(8), 137.

Xu, S. (2023). *Variation in Opportunity to Learn at Secondary Education: The Social Determinants of Between-and Within-School STEM Tracking in the US and Beyond (Doctoral dissertation, University of Pittsburgh)*.

Yang Hansen, K., & Strietholt, R. (2018). Does schooling actually perpetuate educational inequality in mathematics performance? A validity question on the measures of opportunity to learn in PISA. *ZDM: The International Journal on Mathematics Education*, *50*(4), 643–658.