Surfacing Equity Issues in Large Computing Courses with Peer-Ranked, Demographically-Labeled Student Feedback

BENJAMIN XIE, University of Washington, USA
ALANNAH OLESON, University of Washington, USA
JAYNE EVERSON, University of Washington, USA
AMY J. KO, University of Washington, USA

As computing courses become larger, students of minoritized groups continue to disproportionately face challenges that hinder their academic and professional success (e.g. implicit bias, microaggressions, lack of resources, assumptions of preparatory privilege). This can impact career aspirations and sense of belonging in computing communities. Instructors have the power to make immediate changes to support more equitable learning, but they are often unaware of students' challenges. To help both instructors and students understand the inequities in their classes, we developed StudentAmp, an interactive system that uses student feedback and self-reported demographic information (e.g. gender, ethnicity, disability, educational background) to show challenges and how they affect students differently. To help instructors make sense of feedback, StudentAmp ranks challenges by student-perceived disruptiveness. We conducted formative evaluations with five large college computing courses (150 - 750 students) being taught remotely during the COVID-19 pandemic. We found that students shared challenges beyond the scope of the course, perceived sharing information about who they were as useful but potentially dangerous, and that teaching teams were able to use this information to consider the positionality of students sharing challenges. Our findings relate to a central design tension of supporting equity by sharing contextualized information about students while also ensuring their privacy and well-being.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; Interactive systems and tools; • **Social and professional topics** → *Computing education.*

Additional Key Words and Phrases: student feedback, equity, computing education, demographics, learning at scale

ACM Reference Format:
Benjamin Xie, Alannah Oleson, Jayne Everson, and Amy J. Ko. 2022. Surfacing Equity Issues in Large Computing Courses with Peer-Ranked, Demographically-Labeled Student Feedback. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 65 (April 2022), 39 pages. https://doi.org/10.1145/3512912

1 INTRODUCTION: CONTEXTUALIZING FEEDBACK TO UNDERSTAND INEQUITIES

Teaching equitably is important in computer and information sciences (CIS), where there are many inequities in formal CIS courses in university and higher education contexts. Formal higher education in CIS is a primary pathway for participation in the computing community, yet CIS

Authors' addresses: Benjamin Xie, University of Washington, The Information School, DUB Group, Seattle, Washington, 98195, USA, bxie@uw.edu; Alannah Oleson, University of Washington, The Information School, DUB Group, Seattle, Washington, 98195, USA, olesona@uw.edu; Jayne Everson, University of Washington, The Allen School, DUB Group, Seattle, Washington, 98195, USA, everjay@uw.edu; Amy J. Ko, University of Washington, The Information School, DUB Group, Seattle, Washington, 98195, USA, ajko@uw.edu.

2573-0142/2022/4-ART65
https://doi.org/10.1145/3512912

courses face persistent diversity, equity, and inclusion issues [42, 85]. In part due to the growing demand for computing skills in the workforce, CIS enrollment numbers have surged recently, straining the capacity of instructors to scale teaching [61]. Despite their popularity, CIS courses continue to face challenges retaining and supporting diverse students in both high school [24] and college [82, 85, 91]. This results in the loss of diverse potential contributors to computing fields [61] and raises social justice issues around who can access and engage with computing communities [46, 74, 75].

One way to to try to teach more equitably is by sourcing feedback from students and responding to it [16]. Feedback is especially critical to an equitable learning environment because students from minoritized groups face unique challenges that, if left unaddressed, can pose serious impediments to science and technology learning [20]. Instructors in higher education have used student feedback as a way to monitor and improve teaching equality, especially in distance and remote learning environments [43]. However, it is not enough to simply be made aware that these challenges exist: To help turn student feedback into action, instructors need context regarding their students' lived experiences to understand how challenges affect different students [17, 54].

Student feedback tools in large (100-500+ student) courses must be scalable. Prior CSCW works have examined massive open online courses (MOOCs), looking at student motivations and retention [89], the impact of a reputation system on the student experience using forums [13], and how matching students across locations helped students to earn higher grades [48]. But in contrast to MOOCs, remote courses typically have more synchronous interactions and feedback mechanisms between instructors and students. Scalable feedback in synchronous courses involves ensuring convenience for students to share feedback [29, 44], and convenience for instructors and teaching assistants (TAs) to collect, analyze, and discuss the feedback [43, 69]. In addition to this requirement, we argue that student feedback for equity-oriented goals must also provide the context to help instructors and TAs consider feedback within the context of students' lived experiences while also ensuring students' privacy.

Context is important to support equity-oriented goals, but existing student feedback mechanisms lack the context to connect feedback to lived experiences. At the scale of hundreds of students to a single *teaching team* consisting typically of one instructor and a few TAs, a teaching team typically cannot respond to all feedback. As a result, commonly used electronic response systems, such as anonymous online surveys sent to students during the term, often lack context about lived experience [12]. This loss of context results in less actionable feedback as it obscures perspectives of minoritized groups as they become lost amongst the majority of perspectives which are typically from students of dominant groups [54].

Contextualizing feedback can come in tension with protecting student privacy, another critical aspect for equity-oriented feedback. Students from minoritized groups are often most at risk when their information is exposed without their informed consent. Feedback methods that are interpersonal and conversational, such as conversations between a student and an instructor after a lecture or with a teaching assistant (TA) during office hours, are common within large computing courses. While these methods provide context by revealing students' identity, they also privilege students who are more willing and able to speak up, such as white and Asian men with prior programming experience [29]. Interpersonal techniques can be especially problematic for students of minoritized groups due to a lack of anonymity potentially introducing stereotype threat [79, 88] and social-desirability biases [28, 33]. Student feedback for equity-oriented goals must also ensure students' privacy when they share feedback.

To explore the design of a student feedback tool that supports equity-oriented goals by 1) being scalable, 2) providing context, and 3) ensuring student privacy, we designed *StudentAmp*. StudentAmp contextualizes student feedback on challenges in their life with demographics and peer perspectives from other students. Using StudentAmp, students self-report challenges that interfered with their learning as well as demographic information (e.g. gender, ethnicity, prior programming knowledge). Students then consider random pairs of challenges their classmates/peers shared and determine which challenge they would consider more disruptive. StudentAmp aggregates students' *meta-feedback* responses to produce a ranking of perceived challenge disruptiveness. StudentAmp then uses this data to produce a report for an instructor detailing challenges students reported, contextualizing challenges with demographic information as well as ranking them according to student perceptions of disruptiveness.

To evaluate the effects of using StudentAmp to collect and report student feedback contextualized by demographic information and perceptions from classmates, we conducted a formative study with teaching teams and students of five large remote computing courses (163 - 628 students/course). We considered the following research questions:

(1) What do students share about challenges interfering with their learning?
(2) How do students perceive the values and risks of sharing information on challenges they face, contextualized with demographics and peer-perceptions?
(3) How do teaching teams of large computing courses use different types of information to contextualize students challenges for equity-oriented interpretations?

We found that students considered the privacy of themselves and others when sharing feedback on challenges that were often about their lives beyond computing courses. Seeing anonymous peers' challenges also helped students empathize and develop a sense of belonging with peers. Instructors used demographic data to connect challenges to student experiences by situating challenges in lived experiences that may differ from dominant norms, while finding data on peer perspectives questionable and unreliable. We interpreted these findings as design trade-offs between contextualizing feedback with demographic data to inform stakeholders of inequities at scale and ensuring the privacy and well-being of students.

This paper makes the following three contributions:

(1) A large scale thematic analysis of 810 challenges that 604 students faced while learning computing in large remote courses during the COVID-19 pandemic;
(2) An artifact (StudentAmp) that is a design exploration into how contextualizing student feedback may support more equitable learning experiences in large, remote courses; and
(3) A rich qualitative investigation of
    (a) Students' experiences sharing feedback through StudentAmp and viewing the feedback of their peers, and
    (b) Teaching teams' experiences using StudentAmp to better understand inequities in their courses by interpreting contextualized student feedback.

2 BACKGROUND: EQUITY, PERSPECTIVE TAKING, AND THEORY OF ACTION

In this section, we provide our framing of equity within the context of higher education computing courses. Then, we describe how student demographics and perspectives can contextualize student feedback data by providing opportunities for perspective taking and empathizing. Finally, we describe a Theory of Action that scaffolds interpretations and uses of student feedback data.

2.1 Equity Involves Understanding Experiences of Minoritized Groups

We framed equitable learning as ensuring students from diverse backgrounds can successfully access and engage with a learning experience to realize their dignity and potential. Within the context of computing education, understanding students' diverse backgrounds involves considering intersectionality [21], or how different aspects of students' identities intersect and interact. These aspects of identities include students' ethnicities, genders, disabilities (physical, mental, social), preparatory privilege, current situation outside of the course (e.g. familial or financial responsibilities), and past educational experiences (whether they are a transfer student, first generation). Learners are complex individuals beyond a single demographic label.

When considering equity, we must consider not only students, but also the structures of society that students exist in. We establish this framing of equity upon upon Structuration Theory, which defines a recursive process where society and its structures shape the activity of individuals and individuals shape and condition the structures of society [14, 47]. Understanding equity involves not just individuals involved, but also the context of the economic, social, cultural, and political conditions of the time and place [27, 60]. Equity has a social justice goal where corrective measures must adjust for aggregate harm from social inequalities [67]. As a result, understanding equity involves considering how learners are situated within complex environments that they also shape.

Within computing courses, improving equity would require not just improving access to computing education, but also supporting successful participation and achievement by diverse students learning computing [50]. Structural and systemic inequities embedded in and around computing courses can manifest as barriers to participation (e.g. unconscious bias of instructors excluding students of color from successful participation [70]), affect students' sense of belonging and identity (e.g. instructional materials promoting gender bias [57]), and exacerbate existing disparities in privilege (e.g. students cannot synchronously engage with instructors and other classmates because of timezone differences, work commitments, or familial responsibilities) [50]. Inequities arise when structures and norms fail to include or serve students of minoritized groups. Addressing inequities often involves interventions that support the needs of specific groups, such as a one hour social-belonging intervention to support the long term career, mental health, and community building for Black students [7, 84].

For this paper, we referred to *minoritized* as a descriptor of identity groups that are typically not dominant within computing communities in the United States (US). Dominant groups are positively privileged [87], unstigmatized [72], and generally favored by the institutions of society [53], particularly within social, economic, political, and educational systems [23]. For the context of college computer and information science programs in the US, we characterized dominant groups as including white and Asian men who started college shortly after high school (not transfer students), do not have disabilities, have little or no financial or familial responsibilities, have English fluency, and have at least one parent who completed a four year college degree. Minoritized groups, then, are groups that are not positively privileged or favored and often stigmatized. In our context of study, minoritized groups include students who are women, non-binary, African-American/Black, Hispanic/Latinx, Native American/Indigenous, Pacific Islander, transfer students, not fluent in English, and/or first-generation, as well as students who have disabilities and/or have financial or familial responsibilities. While some may consider minoritized groups a small proportion of the population, these groups can actually make up a large proportion of society while still being minoritized by systemic injustices. Systemic cultures and norms tend to favor dominant groups and disadvantage minoritized groups.

2.2 Perspective Taking to Better Understand Students' Situations

To understand others' situations, humans rely (at least partially) on empathy. Empathy is a multidimensional construct, a set of interrelated yet distinct social behaviors and abilities that enable understanding of other peoples' unique contexts [22, 71, 76]. According to some models, there are two major kinds of empathy: affective empathy, which involves responding with one's own emotion to another person's mental or emotional state; and cognitive empathy, which involves the ability to understand another person's mental state [22]. For this investigation, we focused specifically on promoting *perspective-taking* behavior, which is a facet of cognitive empathy involving adopting others' points of view [22].

Perspective taking as a means of empathizing can be useful tool in understanding others' situations and needs. For instance, in design contexts, designers often use some form of perspective taking to try and better understand the needs of different groups of stakeholders. Prior work suggests that encouraging perspective taking through the use of personas or cognitive walkthroughs can help promote better understandings of minoritized groups [32, 59], including people with different genders [10, 80], cultures [1], socioeconomic statuses [58], and abilities [5, 62].

Supporting proper perspective-taking behavior can be challenging, especially in educational contexts. While there remains comparatively little work on promoting empathy in traditional programming courses, prior work in the area of HCI and software design education suggests that perspective taking can be difficult to teach and to learn [32, 63, 64, 68]. This is especially true in higher education computing contexts, which tend to be dominated by young, cognitively and physically high-performing students who may lack exposure to perspectives and viewpoints that are very different than their own [51]. Poorly executed perspective-taking activities may also lead to stereotyping, or making erroneous assumptions about a particular individual based solely on some limited information about them, such as their demographics. Stereotyping is a particular danger when asking people from dominant groups to perspective-take with people in minoritized or less contextually dominant groups [4, 8].

Stereotyping is an innate human behavior and cannot be done away with entirely [83]. For instance, if no particular traits about users are specified, software designers practicing perspective taking may fall back on implicit assumptions that a user is of a contextually dominant race, gender, age, culture, and class, who is heterosexual, affluent, comfortable with technology, and not disabled [19]. However, prior work suggests that providing enough rich contextual information about the target person's identities and behaviors can preclude some of the harmful effects of stereotyping [38]. Providing more information about a person's experiences and identities also can reduce tendencies toward single-axis analysis [19] which can erase the lived experiences of those with intersectional identities. To have the best chances of perspective taking being effective, comprehensive, and beneficial instead of harmful, providing more information about a person can support more holistic understandings of their unique situation.

2.3 Theory of Action to Guide Data Interpretation

Becoming aware of challenges is only a first (but critical) step towards addressing them. That is to say that showing somebody information will not necessarily translate to action. To scaffold this connection between data and action, we used Theory of Action, a framework that helps educators develop evidence-based stories that explains the specific changes they intend to make to improve teaching and learning [11].

We drew upon Theory of Action (ToA) to connect instructors' interpretations of data from StudentAmp to action [11, 15, 41]. ToA relates individual actions to systemic functioning by articulating the underlying logic of work and starting assumptions about how and why actions will lead to desired outcomes [15, 41]. In ToA, actions involve information that stakeholders find valuable within their societal structures and can use to affect power dynamics [15]. While originally derived from studies of individual and organizational learning [2, 3], educational policymakers and administrators have used ToA to make changes to improve teaching and learning [11, 41].

Ongoing development and communication of a ToA can help instructors improve teaching and learning. Most of the work that uses ToA as a guiding framework to improve teaching and learning has primarily focused on schools and school districts teaching primary/elementary and secondary education (e.g. [41, 45]). It is typically an iterative process where leaders look at data to understand students' learning experiences, as well as reflect on how teachers' instruction affects student learning and how school principals' practices affect teachers' instruction [11, 41]. For ToA to make changes that would improve teaching and learning, leaders must articulate ToA and reform plans in terms that are compelling and understandable to multiple stakeholders and lay framework for ongoing "reform conversation" [41].

While leadership in primary and secondary educational institutions in the United Stated tends to be more centralized to school and district levels, post-secondary institutions (e.g. colleges and universities) tend to afford individual teachers (professors) more autonomy over their own classes [9, 35]. Leveraging this, we re-framed ToA to remove the principals and instead have teachers/instructors (e.g. professors, lecturers) as the leaders. This new framework positions instructors within the leadership role of 1) looking at data to understand students' learning experiences, 2) reflecting on how instruction affects student learning, and 3) identifying how the context surrounding the course affects instruction. This context affecting learning is broad and can include departmental policies (e.g. grade being used for acceptance into competitive major) and current events (e.g. global health emergency, political unrest).

To successfully use ToA to improve a course involves having course instructors serve as stewards who continuously develop, communicate, and advocate for actions. Honig et al. 2010 saw stewardship as critical to the ongoing process of reform with ToA [41]. They identified tasks that stewards must take which we adapted from a school/district level to a course level: Ongoing development of a theory of action for the transformation of course; communication with others to help them understand the theory of action, including strategies used and underlying rationale for these strategies; and strategic brokering of external resources and relationships to support the overall course transformation process.

We framed StudentAmp as a tool to provide data that course instructors could use to inform the creation, iteration, and application of Theories of Action to improve their course. We scaffold the implementation of StudentAmp in courses within the context of creating Theories of Action (defined in [11]).

3 DESIGN OF STUDENTAMP

To understand the design of StudentAmp involves first understanding the positionality of the researchers who designed the tool, as well the design considerations we considered. We describe these first, then describe StudentAmp and how we intended for students and instructors to interact with it.

3.1 Critical Self-Reflexivity: Acknowledging Researchers' Positionality

This research required a reduction of people to the responses they were willing to share, so we acknowledge our assumptions and values in this section. By doing so, we follow critical approaches to quantitative methods which require researchers "to engage in critical self-reflexivity as a necessary first step for the long journey of deracializing statistics" [30]. As part of this process, we define assumptions and commitments that were the foundation of this research.

Firstly, we recognize the power structures and heterogeneity of people within different roles. Direct stakeholders in this research included teaching teams (including faculty members leading the teaching of a course and teaching assistants (TAs) supporting the teaching) and students involved in the course. Even within these groups, there were differences. Faculty members leading instruction ranged from tenured research-track faculty who had worked for years at the institution to teaching-track lecturers with comparatively less teaching experience. The TAs were all undergraduate students, but their experiences with their respective courses ranged from never having taught or taken it to having years of previous experience taking and teaching the course. Full or part-time students attending these courses may or may not have been accepted to their major (admissions to CIS majors is very competitive and not guaranteed as part of admission to the university). Most students were enrolled to take the course, but some may have had listener status where they were not taking it for official credit. Some transferred from other higher education institutions (e.g. two-year institutions) with different norms, while others came directly from high school. A common theme across all stakeholders: The data we collect is a partial and biased lens into their experiences in a select few courses as part of a much larger educational experience.

Secondly, we acknowledge the tensions between labeling people in data, the *intersectionality* of people's identities (students in particular), and ensuring privacy. Intersectionality denotes the various ways in which ethnicity and gender (and other demographic labels) interact to shape the peoples' lived experiences [21]. Prior work has found that simplistic labeling of people can harm minoritized groups in particular. Labels of demographics (e.g. ethnicity, gender) academic experience (e.g. year in school, major, transfer or not), and lived experience (e.g. disabilities, familial language) are overly-simplistic. Furthermore, we needed to balance the nuance of the labels we select between how representative they were to diverse individuals and how anonymous they were such that instructors could not map responses back to individuals or small groups of students. Despite these risks, we believed that instructors could still use these labels in such a way to help stakeholders contextualize relationships between challenges and intersectional groups of people. Our perspectives align with the notion that "race is a measure of a relationship – not an inalterable trait" [90].

Finally, we acknowledge that models are always wrong in that they never fully reflect the complex phenomena we want them to represent, but they can be designed such that they are useful in informing stakeholders of hidden challenges. We framed the work we did as producing simplified models of the complex phenomena of inequity in classes. We do not believe that in itself models will help, but they can support conversations, interactions, and interventions that address the systemic issues we sought to bring to light [16]. The objective of this work was to help instructors identify equity issues in their class, and that is a first of many steps in making learning experiences more equitable and just.

3.2 Design Considerations to Support Scalability, Context, and Privacy

We used the following design considerations to guide the design of StudentAmp:

(DC1) **Privacy/anonymity/safety**: Providing feedback should not harm a student. That is, instructors or other students should not be able to map responses back to specific students and information collection should not distress students. Because the anonymity of electronic responses systems can increase students' propensity to engage in providing feedback [29] and low response rates are a common issue with student feedback systems [26, 78], we believe that anonymity will support more inclusive participation.

(DC2) **Potential lack of awareness**: Students are not necessarily aware of all possible challenges they're facing and instructors are not aware of all possible challenges in their classes.

(DC3) **Person-in(fluencing)-environment**: Challenges are artifacts of inadequate support of students from their environment, not inadequacies of individuals.

(DC4) **Time constraints**: Students are limited in their availability and motivation to provide information and teachers are limited in their availability to analyze it. Furthermore, instructors need time to enact changes to their courses.

(DC5) **Relative disruptiveness of challenges**: Some challenges affect a student more or less than other ones.

(DC6) **Proximal, perceived value of participation**: Both students and teachers should perceive tangible and timely value for their participation.
(DC7) **Intersectional perspective of students**: *Intersectionality* denotes the various ways in which race and gender (and other demographic labels) interact to shape the multiple dimensions of underrepresented peoples' experiences race [21]. Whenever possible, StudentAmp should provide insight into the intersectionality and complexity of identity.

These design considerations are not without tensions. In this study, we focused on the tension where designing for equity-oriented goals involved a balance of protecting the privacy of minoritized groups (DC1) while also conveying their intersectional identity (DC7) such that others can better understand their experiences.

3.3 StudentAmp Enables Sharing of Contextualized Student Feedback

We designed StudentAmp as a responsive website to enable broad use by students and teaching staff. In initial interactions with the tool, instructors created their own *sections* to be an instructor of a new course. To grant other users instructor access (e.g. teaching assistants), instructors in the study provided a list of emails to researchers, who then manually gave those accounts access. Students then created accounts by signing up by email or Google account, using a six digit character code provided by their instructor to join a section as a student. Students could join multiple sections, and within the context of this study, we did find that two students were enrolled in multiple courses that used StudentAmp. Users could switch between being students in courses they were enrolled in (via section code) and instructors in courses for which they had instructor permissions, if they had access to any.

3.3.1 Student View: Sharing challenges, demographics, perspectives on other challenges. For this study, we designed StudentAmp's student view with the intention of enabling a student to be able to share feedback within a few minutes. Students could access StudentAmp from any modern web browser (e.g. Firefox, Chrome, Safari).

Figure 1 shows an example of the StudentAmp interface as it appeared to students. Students first shared "the biggest challenge in [their] life getting in the way of this class," with helper text which prompted students to think beyond the scope of the class. Text also appeared which encouraged students to share more ("Keep writing so others understand your challenge!") if their response was < 100 characters and more if their response was ≥ 160 characters ("You wrote quite a bit! Consider condensing your writing so others can read it quickly."). From our pilot testing, we found that a message of 100-160 characters (approximately the maximum length of a tweet on the social media platform *Twitter*) represented sufficient description for another student or instructor to understand a challenge response without being too burdensome to read.

After sharing their challenges, students had the opportunity to self-report demographics, as shown in Figure 1, step 2. We based StudentAmp's demographic questions on factors which prior work found to be impactful to students' learning experiences, including prior programming experience, whether they were a transfer student [49], whether they were first-generation, whether their familial language is the same as the language the course is taught in (English), gender, ethnicity, and physical, mental, or social disability status. While we required answers for all multiple choice demographics questions, each question included an option for "(prefer not to answer)." If students had previously filled out demographic questions (e.g. in a previous feedback session), StudentAmp populated these questions with the student's prior responses. The demographics questions and options were as follows:

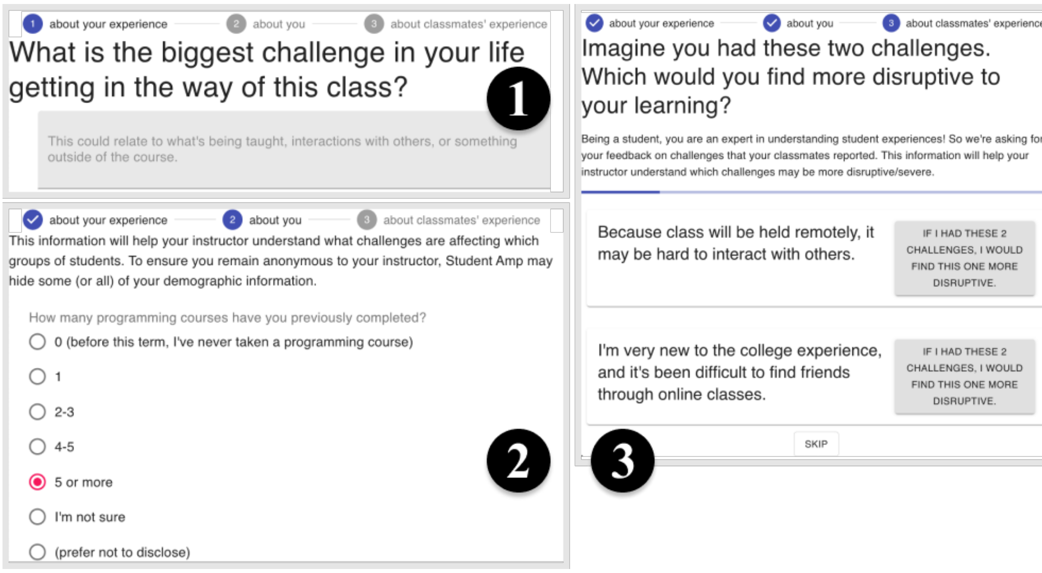

Fig. 1. StudentAmp student view: Students shared 1) a challenge they faced, 2) demographics (pre-populated if they've previously filled in), and finally 3) meta-feedback by selecting which of two random challenges their peers shared was more disruptive, repeating this step two to eight times depending on class size.

(1) *How many programming courses have you previously completed?* 0 (before this term, I've never taken a programming course); 1; 2-3; 4-5; 5 or more; I'm not sure; (prefer not to disclose). [select one]
(2) *Did you previously attend another college/university? (e.g. 2-yr community college, another 4 yr university).* YES, I previously attended another college/university; NO, my current college/university is the first one I have attended; I'm not sure; (prefer not to disclose). [select one]
(3) *Are you a first-generation college student? (first-gen if parent(s) did not complete a 4 yr college/university degree).* YES, I am a first-generation college student; NO, my parent(s) completed a 4 yr college degree; I'm not sure; (prefer not to disclose). [select one]
(4) *Is the language your family primarily speaks at home the same as the one used to teach this class?* YES, the language my family primarily speaks at home is the same as the one used to teach this class; NO, my family speaks a different language than the one used to teach this class; I'm not sure; (prefer not to disclose). [select one]
(5) *Are you currently working or searching for a job? (select all that apply).* I am actively looking for a job; I work part-time (20 hrs a week or less); I work full-time (more than 20 hrs a week); I am neither working nor looking for a job; I'm not sure; (prefer not to disclose). [select one or more]
(6) *What is your gender? (select all that apply).* woman; man; non-binary; prefer to self-describe**; (prefer not to disclose). [select one or more]

(7) *What is your ethnicity? (select all that apply).* Asian; Black/African; Hispanic/Latinx; Native American; Pacific Islander; white; prefer to self-describe[1]; (prefer not to disclose). [select one or more]
(8) *Rate to what extent a physical/bodily disorder hinders your learning experience.* 0: not at all; 1: to a small extent; 2: to some extent; 3: to a moderate extent; 4: to a great extent; 5: to a very great extent; I'm not sure; (prefer not to disclose). [select one]
(9) *Rate to what extent a mental or social disorder hinders your learning experience.* 0: not at all; 1: to a small extent; 2: to some extent; 3: to a moderate extent; 4: to a great extent; 5: to a very great extent; I'm not sure; (prefer not to disclose). [select one]

After sharing their challenges and demographics, students finally shared *meta-feedback* on their classmates' responses, as shown in Figure 1.3. In this phase, StudentAmp randomly selected two challenges classmates had reported and asked students to imagine they had these two challenges, then to select the one that they imagined would be more disruptive to learning. To support data integrity of the meta-feedback, students could *skip* any responses. The meta-feedback pairwise comparison process was repeated 2-8 times depending on the class size ($2 \leq 2 * log(class\ size) \leq 8$).

3.3.2 Instructor View: Designing for instructor-led data exploration. Once students shared feedback, StudentAmp presented teaching teams with a report on student feedback, as shown in Figure 2. We designed StudentAmp to augment instructors' domain knowledge related to the course and their students by enabling exploration of contextualized feedback data. StudentAmp enabled this data exploration by 1) informing instructors of challenges, which student groups they affected, and how severe students perceived them to be and 2) supporting situated annotations through labels and notes so instructors could review previous findings.

As mentioned previously, instructors within our study each created a *section* for their course which students joined via a unique 6 character code. Each time an instructor wished to use StudentAmp to gather feedback, they created new feedback session. Once students shared feedback via StudentAmp instructors viewed the results, as shown in Figure 2. Teaching teams could review this data to identify how certain types of challenges disproportionately affected certain groups. Instructors and teaching teams could browse reported challenges, sorted by disrupt score (Fig. 2e), identifying trends and patterns. They could then create labels and assign them to challenges (Fig. 2f). We designed StudentAmp's labels to help teaching teams organize and prioritize feedback to better identify trends within and across feedback sessions. Similar to GitHub labels [31], teaching teams could define any labels they wanted, then assign one or many labels to any responses in any feedback session, similar to a tagging system. While the use of labels did require teaching manually labeling individual feedback (e.g. we did not provide automated labeling), it also enabled filtering of responses to use demographics charts (Fig. 2c) to explore how challenges affected demographic groups. In our study, we saw that this helped teaching teams understand which types of challenges disproportionately affected different groups of students across the nine demographic features we collected (enumerated in section 3.3.1).

In StudentAmp, we aimed to support beneficial and effective perspective taking for instructors. Teaching teams could look at challenges (unfiltered or filtered by label) and use demographic information associated with individual challenges (Fig. 2d) to perspective take. To help instructors better understand the nuanced ways that different challenges affected different students, we provided demographic information alongside each challenge to promote more informed perspective taking through the addition of richer contextual information. By doing so, we hoped to avoid

[1] To support more inclusive demographics reporting [77], a free response follow-up question appeared with the prompt "Please self-describe your {gender, ethnicity}" after a student selected "prefer to self-describe" for gender or ethnicity. This information was not shared with teaching teams due to privacy concerns.

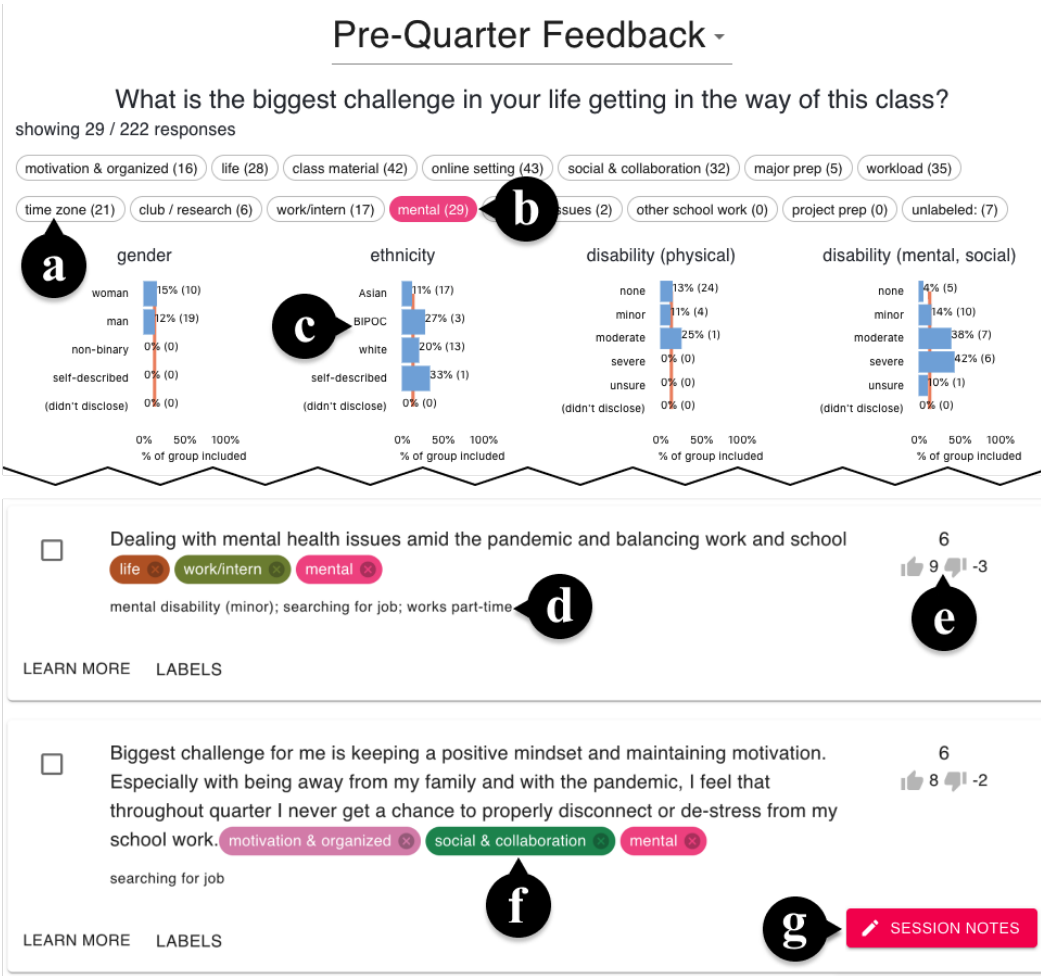

Fig. 2. StudentAmp instructor view: Teaching teams could organize challenges by creating custom labels (a), which they could select to filter responses (b). The filters enabled teaching teams to use charts of demographic information (c) to see how challenges disproportionately affected certain groups (e.g. how the 29 challenges labels "mental" disproportionately affected BIPOC students and students with moderate or severe disabilities. The instructor view also included each challenge that included the selected label(s). Each challenge was contextualized with demographics for minoritized groups that students identified with (d), disrupt score (e), and labels that the teaching team assigned to that challenge (f). Teaching teams could also share collaborative notes (g), which have prompting based on our Theory of Action.

stereotyping by encouraging instructors to see their students as unique individuals with many different kinds of identities and contexts, rather than defaulting to the assumption of an "average" (likely dominant) student if no information was given. However, we also had to balance this against the need to protect student privacy and preserve anonymity, which we discuss further in our design considerations.

Throughout this entire process, teaching teams could use StudentAmp's collaborative *notes* (Fig. 2g) to review Theory of Action principles and also share notes with other members of the teaching team. Notes were open-ended text fields shared by all teaching teams. Each feedback session had a separate notes section from which instructors could write notes about their findings within the report. To help guide teaching teams according to our Theory of Action, we included prompting above the notes. This prompting asked teaching teams to consider 1) What's going on in students' learning experiences?, 2) How does what instructors are doing (or not doing) affect learning experiences? and 3) What factors external to the course help or hinder students' learning experiences? Which students? It then asks students to fill in the blanks as many times as possible: "If I/we __, then the course will change by __, so that students who are __ will be able to __."

3.3.3 Design of Disrupt Score. We designed StudentAmp's *disrupt score* with an intention to draw attention to challenges that were more disruptive to students' learning, rather than those that were simply more frequent or relatable to the majority of students. Prior work [20] and our pilot testing identified that students of minoritized groups reported challenges that were often unique from what their peers shared. As a result, we designed StudentAmp to support equitable feedback processes by organizing and presenting feedback in a way that went beyond showing the most common challenges. The disrupt score provided a quantification metric intended to represent how disruptive to students' learning some challenges were compared to other challenges.

In its most literal sense, the disrupt score was the net number of times a random student within the course decided a challenge was more disruptive to learning than another randomly selected challenge. It was based on a pairwise comparison of random challenges, as shown in Figure 1.3 (see also the description of *meta-feedback* above). We based this strategy on the Copeland method of pairwise scoring to determine ranked voting [18, 73]. Each challenge began with a score of 0. Each time a student shared meta-feedback (described in 3.3.1), the challenge they selected had its disrupt score incremented by one (+1). The challenge they did not select correspondingly had its disrupt score decremented by one(−1). No score changes occurred if a student chose to skip a meta-feedback comparison. These scores were then aggregated and used to track challenges in the instructor view in addition to being shown alongside each challenge (Fig. 2e).

A key assumption to the disrupt score was that students were able to perspective take and consider the disruptiveness of challenges they may not even have. Our initial pilot testing with students thinking aloud as they considered challenges found that students were more likely to select the challenge they most associated with having. As a result, for the study results we report in this paper, we adjusted StudentAmp's design by adding more information in prompts and button text to explicitly encourage consideration of disruptiveness rather than readability. As we report and discuss in the following sections, think-alouds with interviewed students suggested that students conceptualized the meta-feedback process in different ways (section 5.2.4), which resulted in several teaching teams having trouble interpreting disrupt scores (section 5.3.3).

4 STUDY DESIGN: DEPLOYED STUDENTAMP AND CONDUCTED INTERVIEWS

4.1 Context: Five large, remote, computing courses during a pandemic

To answer our research questions, we conducted a formative evaluation where five large computing courses used StudentAmp over the course of an 11 week term and we surveyed and interviewed professors, TAs, and some students twice throughout the term. An institutional review board (IRB) approved this study prior to any data collection.

We recruited culturally competent [86] instructors teaching large computing courses (100+ students). We chose to directly contact instructors who had previously demonstrated cultural competence [86] for participation in the study, as evidenced by their prior research efforts or participation in seminars on anti-racism. We focused on culturally-competent professors to avoid interpretations of student feedback and demographics focused on innate ability, such as certain people or groups having the "geek gene" and being more suited to computing [55, 65]. We selected large courses to better ensure anonymity and to have a scale at which analyzing individual feedback of an entire course would become time-consuming and challenging. To avoid potentially unproductive and harmful uses of StudentAmp, we chose to exclude instructors who had previously stated beliefs about demographic groups' interest and abilities in computing [65].

Of the eight professors we reached out to, five professors teaching large computing courses participated in this study and used StudentAmp. Table 1 provides an overview of those five courses and how they used StudentAmp. Courses all took place within the same term, during which a global COVID-19 pandemic and a global reckoning with racial injustice were both ongoing issues. Because the university was shut down to in-person learning, all courses in this study were taught remotely, with students located in time zones all over the world. Given courses B, D, and E were introductory courses with no prerequisite requirements, some students in the study were first year college students who had yet to have an in-person college experience.

The five courses in our study were from the two departments of the same research university. This public research university in the United States was located in a major city with significant presence of large technology companies. All participants from this study (professors, TAs, and students) taught at or attended this university. We interpreted this study context to be Western, Educated, Industrialized, Rich, and Demographic (a.k.a. WEIRD, [37]). While most people around the world are not from WEIRD societies [36], many computing education contexts in the United States tend to be WEIRD societies [82, 91].

Courses A, B, and C were different courses offered within the computer science (CS) department. Based on data collected earlier in the same academic year that this study occurred, the department has 1,668 enrolled undergraduates. The CS department reported 31% of their students as female and 69% of undergraduates as male (only binary gender was collected on this survey); 8% of students as under-represented minority/URM (African American, American Indian/Alaska Native, Hawaiian/Pacific Islander and Latinx/Hispanic) and 75% as non-URM; 20% as first-generation (none of their parents completed four year college degrees); and 17% as international.

Courses D and E were the same course taught by different teaching teams and from the information science department, a separate department from the CS department. Based on enrollment data collected the term after this study, the information science department had 526 undergraduates. The information science department reported 43% of undergraduates as female and 57% as male (only binary gender collected). The department reported 45% as Asian, 12% as URM, 16% as white, 28% unknown. Whether a student was Hispanic/Latinx was reported separately, with 4% (23) reporting as Hispanic/Latinx.

In addition to using StudentAmp, teaching teams relied on other tools and methods to collect student feedback for various purposes. Other feedback tools included feedback after assignments (e.g. to find out how long an assignment took, P-B), during lecture (e.g. "to ascertain skill acquisition," P-D), during the middle of the quarter (e.g. mid-term feedback conducted with an instructional consultant, P-A, P-C), and at the end of the term. They had previously used feedback to "respond to small conveniences that students requested" (P-C). They also invited students to reach out to them directly through email or similar mediums, but P-E noted how students from dominant groups tended to speak up more through these channels:

> P-E: *"students that have taken a bunch of programming classes and already done this stuff...they're the ones who speak up, who talk, engage [...] and this is all tied to race and gender."*

4.1.1 Teaching team demographics: Professors from Dominant Groups, TAs were gender-diverse. The teaching teams were led by five professors who were white or Asian men with prior teaching

Table 1. Context about five courses in study and their Student Amp usage: Course content and structure, professors' definitions of equity, number of students who completed course, number of students, number of responses in each StAmp feedback session (number of incomplete responses in parentheses), and who amongst the teaching team had Student Amp access.

ID	Course Content	Course Structure	Prof's dfn. of equity	Students	Responses	Access
A	Intro. to CS II. Data structures, complexity, sorting in Java. One prerequisite course.	optional, recorded lectures w/ professor (3 / wk) and lab section w/ TA (1 / wk)	"everybody should be able to succeed... my focus has been to remove as many structural barriers within the course"	500-750	wk 2: 148 (+3) wk 4: 139 (+7) wk 7: 86 (+8)	professor & lead TAs (6)
B	Intro to CS for non-majors. Control & data abstraction, file processing, visualization in Python.	optional, recorded lectures w/ professor (3 / wk) and lab section w/ TA (1 / wk)	"there are a lot of cultural problems in the CS space... elitism and racism, to some degree, and sexism. It was an important to me that I can try to address those impressions"	150-200	wk 0,1: 30(+8) wk 2: (8) wk 3: (1) wk 4: (2) wk 7: (4)	professor only (all 7 TAs saw responses)
C	Design, analysis, and critique of data structures and algorithms in Java. Course A is prerequisite.	Students meet in small groups w/ TA 4 days / wk. Assignments: three group projects, each two weeks long.	"How are we engaging with students' identity in the course"	250-300	wk 0,1: 218(+4) wk 4: 19 (+10)	professor & head TAs (3)
D	Introduction to collection, storage, analysis, and visualization of data in R.	optional, recorded lectures w/ professor (2 / wk) and lab section w/ TA (1 / wk)	"students should have equal probabilities of success regardless of their background... putting forth the support and resources necessary to balance out the playing field "	150-200	wk 1: 58 (+4) wk 4: 35 (+5)	professor & all TAs (10)
E	(same as D)	(same as D)	"[students are] all able to get to the same ending objective... put the most resources that I have (time and energy) towards supporting [students with the farthest to go]"	150-200	wk 2: (8) wk 5: 13 (+7)	professor only

experience. Professors of all five courses reported as white or Asian men with no physical, mental, or social disabilities. Two professors (for courses A and B) were teaching their courses for the first time but had experience teaching related courses; the other three had taught that same course multiple times before. All five professors had been teaching courses remotely for at least the two terms prior to the study.

Of the 26 teaching assistants (TAs) who had access to or reviewed StudentAmp responses, 17 responded to a survey to report their their demographics. For ethnicity, 1 TA reported as Hispanic/Latinx, 12 as Asian, and 4 as white (non-Hispanic). For gender, 1 reported as non-binary, 9 as women, and 8 as men. Three TAs reported mental or social disabilities, such as anxiety. All but one TA who responded had previously either taken the course or an equivalent one to the course they were serving as a TA for. That one TA who had no prior experience as a student or TA for the course they were teaching had previously served as TA for the course B professor (P-B) for other courses in the past. Taken together, we can say that TAs were predominantly white or Asian, identified as a diversity of genders, and generally had prior experience with the course material as a student and/or a TA.

4.2 Data: StudentAmp responses, interviews w/ students & teaching teams

To answer our three research questions, we collected data from Student Amp and conducted two rounds of individual interviews with students and group interviews with teaching teams.

4.2.1 Data for RQ1: 810 challenges shared with StudentAmp. To understand what challenged students shared, we analyzed challenges shared by students with StudentAmp. In total, 604 unique students shared a total of 810 challenges across the five courses through StudentAmp over the duration of the 11 week term. The *Responses* column in Table 1 shows the number of responses in the feedback sessions across the five courses that used StudentAmp. We included incomplete responses because those responses were only incomplete because those students did not provide meta-feedback.

4.2.2 Data for RQ2: Rounds of interviews with 5 students of minoritized groups to understand perceptions. To understand factors that may have impacted what students shared with their instructors through StudentAmp, we conducted two rounds of semi-structured interviews with students. We recruited students who used StudentAmp and indicated interest in conducting follow-up interviews for compensation ($50 for two 1-1.5 hr interviews, which was slightly above minimum wage in the area at the time). We conducted interviews remotely, recording video (including screen share) and audio with the consent of students.

We interviewed five students from minoritized groups. Of the 234 students who indicated potential interest in a follow-up interview, we identified 39 who were from minoritized groups (as evidenced by their ethnicity, gender) and/or reported a unique challenge. We contacted those 39 students by email and ended up interviewing all five students who replied. These five students came from three courses (three from course A, one from C, one from D). Three students were Asian women, one was a Hispanic/Latinx woman, and one was a white non-binary person[2]. Two were third year undergraduates studying majors related to computer and information sciences (CIS), two were first year undergraduates interested (but not yet enrolled) in CIS majors, and one was second year Master's student studying information science. While taking this course, interviewed students reported other commitments including submitting more than 100 job applications, moving physical locations, having familial responsibilities, and taking almost double the recommended course load.

We collected data via two rounds of retrospective think-aloud style interviews [25] to understand how students interpreted the prompts and how they decided to share what they did.

[2]Regardless of their reported gender identity, we choose to refer to all interviewed students, professors, and TAs using *they/them* pronouns in this paper to discourage re-identification.

The first round of interviews occurred within the first five weeks of the term, after students had shared feedback using StudentAmp at least once. Students answered questions about their experiences learning computing and expectations about using StudentAmp, and then shared their screen as they walked through their prior usage of StudentAmp (three steps in Fig. 1), reviewing and reflecting on their previous responses and the context surrounding them. We asked them to interpret what each page was asking them to do and also about their perceived risks and benefits for sharing information at each page. We closed the first interview by asking students to reflect.

The second round of interviews occurred during the final two weeks of the term and included a sorting activity to have students consider additional challenges they did not share as well feedback on an example instructor view (Fig. 2). These interviews asked students to reflect on their course experience. We then had students look at all the challenges they shared over the course of term (2-3 challenges per students), asking them to identify potential trends, explain how they decided to share these challenges, and consider how peers potentially seeing their reported challenges affected their decisions to share.We then showed students the StudentAmp instructor view to get their perspectives on the utility and risks of contextualized feedback.

4.2.3 Data for RQ3: Rounds of interviews with 5 teaching teams to understand use of StudentAmp. To understand how teaching teams used contextualized student feedback for equity-oriented interpretation, we conducted semi-structured interviews with teaching teams who used StudentAmp.

Prior to or during the first two weeks of the course, we helped members of the courses' teaching teams (professors and TAs) set up their StudentAmp accounts online so they could collect and analyze feedback from their students. We also asked professors to fill out a survey to share their prior experiences with student feedback, the course they were teaching, and demographic information. TAs filled out a similar survey at the end of the term. To understand how teaching teams of large computing courses used different types of information to contextualize students challenges for equity-oriented interpretations, we conducted three interviews with members of the teaching teams of the five courses in our study: One interview with professors individually, followed by two group interviews with the entire teaching team of a course. The interviews with individual professors took place either before the term began or within the first four weeks of the course. In these interviews, we asked professors to share about the course they were teaching, how they intended to collect feedback from students about various aspects of the course (whether that method was StudentAmp or not), and about their personal definitions of equity in educational spaces (see Table 1).

The two group interviews were spaced out across the duration of the course, with one occurring towards the beginning of the term (after at least one feedback session with StudentAmp had been completed) and one at the end. We began the first group interviews by explaining the features of the StudentAmp interface (Fig. 2c,d,e,g), allowing time for the teaching team to ask questions as needed. We then engaged in a process following our Theory of Action (Section 2.3) in which the teaching team tried to understand students' learning experiences using StudentAmp, considered how their teaching practices or external factors might be affecting student learning, and then proposing changes to the course and articulating how those changes might affect students of different groups, identities, or experiences. We asked one member of the teaching team to share their screen during this process, so participants could have a single shared view to discuss. After ten minutes of unstructured exploration of students' responses, if teaching teams were not already doing so, we prompted them to begin identifying broader patterns and trends they saw among student responses, create labels (Fig. 2a), and add labels to challenges (Fig 2f). After a few more minutes of labeling (the exact duration of which depended on each teaching team's level of discussion), we prompted teams to consider the demographic information provided at the top of the StudentAmp interface (Fig 2c) and explore any potential interactions of demographics with the labels they had created to explore how groups of similar challenges might disproportionately affect certain groups of students. The second group interviews followed a similar process, with the main difference being that the student responses under review were from subsequent StudentAmp feedback sessions.

Interviews were conducted remotely with the Zoom video conferencing tool, which enabled synchronous video and audio conversation, messaging, and recording. We choose to use this tool because all five teaching teams were familiar with it as they used it in their remote teaching. We recorded video and audio for all interviews. All members of the teaching team consented to the inclusion of their audio and visual recordings, as well as their survey data in this study. Participants were also compensated at a rate of $15/hr (approximately minimum wage for the area).

5 ANALYSIS & RESULTS: ANALYZING STUDENTAMP RESPONSES, INTERVIEWS

In the following sections, when we present quotes and other data tied directly to an individual, we gave them a unique anonymous ID associated with their role and course. These three part IDs take the form of *<role>* - *<course>* - *<optional number>*, where role is a character that denotes the individual's role within the class (S-Student, P-Professor, T-Teaching assistant), course is the character corresponding to the individual's associated course from Table 1 (A, B, C, D, or E), and number denotes an individual student within the course (from 1 to course enrollment) or an individual TA (from 1 to the maximum number of course TAs). Professors do not have numbers attached to their IDs (e.g. P-A, P-B).

5.1 RQ1: Students shared challenges beyond the scope of the course

In total, 604 unique students shared a total of 810 challenges across the five courses through StudentAmp over the duration of the 11 week term. To better understand what students shared, we conducted an inductive thematic analysis and a subsequent round of qualitative coding using themes from the initial analysis. Three researchers participated in the qualitative analysis:

- The first author, a critical data studies and computing education researcher with seven years of research experience in data equity in computing education. The first author had expertise in designing interactions with data in educational contexts and mixed methods, having previously taught high school and college courses on introductory computer science and data science. He led the design of StudentAmp and the evaluation.
- The second author, a computing education researcher with seven years of research experience in HCI and design methods, including four years researching computing education within that space and a year of teaching experience in higher education computing contexts. The second author had expertise in qualitative methods and led the analysis of student-reported challenges.
- The third author, a computing education researcher with eleven years of middle and high school teaching experience, and two years of educational research in computing-related contexts.

First, all three researchers participated in collaborative affinity diagramming of 100 randomly selected challenges to inductively generate initial themes with a sensitizing concept ([6, 66]) of *types of challenges*. We used these themes as the basis of our code set. All three researchers collaboratively coded 40 (5%) randomly sampled challenges with the initial code set, discussing discrepancies and iteratively refining the code set and code definitions as needed. As can be seen in many of the below quotes, even though we asked students to report the single biggest challenge they were facing in the class, students often reported multiple, often interwoven, challenges. As a result, we allowed for multiple codes per challenge during the qualitative coding effort. Our goal with this coding effort was to achieve consensus, so all codes applied to a challenge had to be agreed upon by all coders.

After that, two researchers (the second and third authors) continued collaboratively coding another 160 (15%) challenges to ensure both of them had similar interpretations of the code set definitions and to address any confusions that arose. Finally, the two researchers divided the rest of the data and each qualitatively coded half of the remaining student-reported challenges. Once they finished their respective analyses, the two researchers asynchronously verified each others' code applications, marking any instances of disagreement. Finally, the two researchers met synchronously to discuss and come to consensus on the codes applied to the final 610 student-reported challenges, discussing interpretations and eventually achieving full agreement on all codes.

Table 2 shows our code set, comprised of the major themes that arose from our analysis and one "Other" code that was applied when a challenge was too unclear to code or when it otherwise did not fall into a coded category. These 16 types of challenges in Table 2 represent different learning difficulties that students conveyed to instructors through StudentAmp. For this analysis, we adhere to Hammer and Berland's perspective on qualitative coding [34], treating the results of our coding effort as organizations of claims about data rather than quantitative data in and of itself. As a result, we do not report specific code frequencies, instead focusing on representative descriptions of the themes observed within our data. In the following section, IDs preceding quotes indicate the speaker's role within the class (S for student), the course in which the student was enrolled (see Table 1), and a randomly generated number unique to each student within the course.

Even though StudentAmp's instructor view shows student-reported challenges alongside some demographic information about the student who wrote it, for the purposes of this paper, we choose not to report demographic information of the speaker for each individual quote. Instead, we report the demographics or our participants in aggregate, to illustrate the diversity of perspectives and experiences represented by the data while still preserving our participants' anonymity and reducing the risk of community or peer re-identification. The following subsection contains quotes from 21 different students to illustrate the kinds of challenges students reported through StudentAmp. 20 of the 21 students who provided these quotes identified as belonging to at least one minoritized group, and often several. Of these 21 students[3], at the time of the study:

- 10 identified as women, 10 as men, and 1 declined to provide gender information.
- 12 identified as Asian, 7 as white, 2 as Hispanic/Latinx, 1 as Black/African, and 1 as Pacific Islander.
- 3 reported taking their first programming course.
- 4 reported attending another institution prior to their current one (transfer students).
- 5 reported as being first-generation college students.
- 6 reported that their family spoke a language at home than was different from the one used in the course (English).
- 6 students reported that they were currently working part-time, and 2 students full-time. 10 students were actively job-searching (e.g. applying to jobs, attending interviews). 5 students were neither working nor job-searching.
- 12 students reported that they did not have a physical/bodily disorder that hindered their learning experience (0 on a scale of 0-5). 8 students reported that they had physical/bodily disorders which hindered their learning to a minor extent (1-2 on scale), and 1 to a severe extent (4-5 on scale).
- 12 students reported that they did not have a mental or social disorder that hindered their learning experience (0 on a scale of 0-5). 2 students reported that they had mental or social

[3]Numbers reported for demographic facets may total more than 21, since students could belong to multiple categories simultaneously (e.g. holding more than one ethnic identity, or both working full-time *and* actively job-searching).

Table 2. Types of challenges that students reported through Student Amp, used as the codeset for our RQ1 analysis. The rightmost columns indicate courses in which at least one student reported an instance of that challenge. The leftmost columns represent categories of themes which arose during our analysis.

				Courses with 1+ reported (Total num. challenges)				
		Code	**Definition: Challenges related to...**	**A (386)**	**B (46)**	**C (253)**	**D (104)**	**E (31)**
	Course-related feedback	*Course structure*	The "how" of the course: assignments, materials, speed, tools and platforms, teaching methods, office hours, etc.	✓	✓	✓	✓	✓
		Course content	The "what" of the course: Topics of instruction, such as computing, math, or programming	✓	✓	✓	✓	✓
		Remote learning	Online instruction methods and tools, not taking the course in-person	✓	✓	✓	✓	✓
External responsibilities, roles, and contexts	Broader academic life	*Other classes*	Workload or time constraints from taking other courses concurrently	✓	✓	✓	✓	✓
		Extracurriculars	Student life-related activities outside the scope of courses, such as clubs, sports, etc.	✓	✓	✓	✓	✓
		Academic context	Departmental or university-wide academic landscape, such as the highly competitive student climate, changing majors, etc.	✓		✓	✓	✓
	Non-academic roles	*Home & family*	Household or familial responsibilities, including roommates, partners, and other relationships	✓	✓	✓	✓	✓
		Job	Work and internship-related activities, including job searching	✓	✓	✓	✓	✓
	Environment and context	*Location*	Geographic location, especially that which differs from the university	✓	✓	✓	✓	✓
		Political	Politically or nationally relevant events, contexts, and/or climates	✓		✓		
		COVID-19	Explicit mentions of the COVID-19 global pandemic, quarantine, lockdown, etc.	✓	✓	✓	✓	✓
Well-being, health, and individual challenges	Well-being	*Physical health*	Physical injuries, bodily wellness, exercise, nutrition, sleep, etc.	✓		✓	✓	
		Mental health	Anxiety, depression, pressure and/or stress, etc.	✓	✓	✓	✓	✓
		Isolation	Being alone or lonely, including difficulties making friends in a course or connecting with others	✓	✓	✓	✓	✓
	Self-regulation	*Motivation*	Ability to focus on and finish a task, including references to procrastination and perceived lack of productivity	✓	✓	✓	✓	✓
		Time management	Ability to balance many competing responsibilities from classwork, jobs, family, personal lives, etc. within time constraints	✓	✓	✓	✓	✓
		Other	Challenges that were listed as "N/A" or "nothing", or that did not contain sufficient data to interpret	✓	✓	✓	✓	

disorders which hindered their learning to a minor extent (1-2 on scale), 5 to a moderate extent (3 on scale), and 3 to a severe extent (4-5 on scale).

The quotes presented below have been edited as little as possible to preserve authenticity. When clarifications or minor edits for anonymity were necessary, or when some less relevant parts of the quote were removed for length reasons, we designate any edits with square brackets.

We found the types of challenges students reported to fall into six broad categories (see the second-from-the-left column of Table 2 for an overview).

5.1.1 Course-related feedback focused on course & remote learning. The first category was that of feedback related to the course itself, represented by the *Course structure*, *Course content*, and *Remote learning* codes. Students who reported challenges with *Course structure* codes often wrote about their difficulties keeping up with the pacing of the course:

> S-A-102: *"The structure of this class because we simultaneously learn stuff for the assessment while learning the stuff for the following week forcing me to sacrifice one for the other."*

Other students who reported *Course structure* codes faced challenges with managing the course's required virtual learning tools or adapting to the instructor's pedagogical style.

When students reported challenges that contained *Course content* codes, they often mentioned the stress that came from trying to learn computing topics:

> S-E-28: *"I have never learned coding/data analysis ever in my life, things are just intimidating. IDK this class is STRESSING ME OUT."*

Many students who reported this challenge also reported not having much prior experience with computing, or who hadn't programmed in a long time.

Remote learning challenges were often reported by students who disliked the virtual format of classes, which was mandated by the university in response to the ongoing COVID-19 pandemic:

> S-E-6: *"The biggest challenge is just the general lack of structure that is inherent in online classes, regardless of how well the instructor organizes the course."*

Some students felt that virtual classes were not as conducive to learning as in-person classes, or that they did not feel like they got as much out of remote classes:

> S-C-109: *"I think the biggest challenge will truly just be the online nature of life right now. Screen fatigue is a big issue for me, & especially knowing that a programming class will require large amounts of screen time after class is a bit daunting. [...] I am worried about feeling intellectually gratified just because of Zoom fatigue."*

Other students who reported challenges containing *Remote learning* mentioned issues with poor Internet connections that made it challenging for them to attend virtual classes and difficulties connecting with peers and teaching staff.

Overall, challenges coded as *Course structure*, *Course content*, and *Remote learning* codes were likely similar to the kinds of feedback instructors might get with traditional feedback mechanisms (surveys, teaching evaluations, etc.).

5.1.2 Broader academic life focused on academic commitments beyond the course. A second higher-level category students reported through StudentAmp was that of challenges in their academic life outside of that particular course, represented by the *Other classes*, *Extracurriculars*, and *Academic context* codes.

Students who wrote about *Other classes* challenges often described the heavy course load they were taking alongside the course in which StudentAmp was used, forcing them to have to prioritize what work they did. Several students wrote about feeling overwhelmed by their academic workload:

> S-D-57: *"my course workload for my other classes is very heavy and my life is being consumed with all of my classes"*

Similarly, some students reported that being involved in various *Extracurriculars* impacted their available time:

> S-D-38: *"This quarter I am doing a few too many club activities and thus it's making it difficult to focus on my classes. It's my own fault."*

The university at which the study took place has a strong culture of student extracurricular involvement, in part due to the fiercely competitive climate within the university's computing-related departments.

Challenges having to do with the particular departmental or university-wide climate were reported by students in challenges involving *Academic Context* codes, such as those that described self-comparison to peers within the computing major:

> S-A-21: *"comparing myself to others; imposter syndrome; competitive environment in computing majors at [university]"*

The competitive, closed major system of the university, in which students were not guaranteed to get into their first choice of major, also contributed to students' stress and was listed as a common challenge due to the timing of the study, which occurred during major application cycles.

These three types of challenges represented by difficulties students were facing that still had to do with their academic lives, but that were explicitly outside the scope of the course itself.

5.1.3 Non-academic roles include familial and job commitments. A third higher-level theme of challenges which surfaced during our analysis was that of non-academic roles and responsibilities, represented by the *Home & family* and *Job* codes.

Students who reported *Home & family* challenges often described difficulties focusing on coursework in their current environments, which often co-occurred with *Remote learning* codes. Sometimes, students simply mentioned that it was difficult for them to focus in their home environments, leading to them having to re-watch lectures or take extra time reading course materials. Other students wrote about their roles as caretakers of other family members, which took time away from their own responsibilities:

> S-B-18: *"I have a sibling that is disabled and another sibling that just started kindergarten. Because of this, I have to help my parents with making sure they attend their classes and do their homework, which is time consuming and also takes time out of other responsibilities around the house, plus work."*

Other household stressors, such as sick pets or siblings, also affected students' abilities to focus, whether due to stress about their well-being or having to provide transportation to medical appointments.

> S-E-27: *"My dog is getting eye surgery today, because of suspected cancer that caused Gluacoma. We don't know if the cancer is malignant yet but it's hard knowing that each day could be his last. :(so far our luck has been pretty bad but I really hope that the other tumors are benign like the one on his stomach. [...]"*

Many students worked jobs or internships during the quarter, or were actively job searching, as represented by the *Job* code. Often, this challenge was discussed in terms of time constraints, which sometimes made it difficult for students to engage with instruction or find time to complete their assigned work. Several students mentioned they were working full-time or part-time jobs alongside their full-time course loads. Other students described their roles as primary providers for their families:

> S-A-71: *"I am the only one supporting my family economically, so they depend on me working and getting money for our dependancies."*

Sometimes, the job environment or tasks themselves contributed to overall student stress:

> S-B-29: *"My job. I work at a homeless shelter. I only work two shifts a week, but I deal with a lot of very high stress situations (fights, 911 calls, suicidal ideation, sexual assaults, mental health crises, etc). Balancing school stress and job stress can often be difficult. It has been harder recently as I am regularly exposed to Covid positive individuals, so the likelihood of me catching Covid is very high."*

Both *Home & family* and *Job* codes represent roles and responsibilities in students' broader lives that placed demands upon their time and available physical, mental, and emotional resources.

5.1.4 Environment and context focus on broader contexts during a pandemic. Sometimes students reported challenges that had to do with their broader contexts, such as those that were classified as *Location*, *Political*, or *COVID-19* codes.

By far the most commonly mentioned challenge within the *Location* code was that of being in a different time zone than the university (likely due to remote learning mandates), making it difficult to attend synchronous classes, work with group members on class projects, and attend office hours.

> S-C-189: *"I think is time. I currently living in [other country] so that I need to get up at 5 o'clock to have this class."*

Students who described challenges relating to *Location* codes sometimes mentioned the weather in the place they were located impacting their mood, and thus ability to learn, as well.

Political codes were somewhat rare, but seemed to strongly impact students when they arose. This study took place at a U.S.-based university at a time when nationally relevant events were regularly occurring, which was stressful and distracting for students.

> S-C-117: *"It is hard to focus on the course during the global pandemic and political instability. I'm very distracted."*

Similarly, many students reported the ongoing COVID-19 pandemic as the greatest challenge detracting form their learning, as represented by the *COVID-19* code. Several students simply wrote some variation on "COVID" or "the pandemic" as their response. Others described the impact the long-term stress and lockdown conditions had on their ability to learn and complete coursework, especially in the context of remote learning. Some students had family members or close friends that had contracted or were recovering from COVID-19 as well, which caused worry and extra stress:

> S-C-124: *"I guess the biggest challenge right now is the health of my family (relatives) since a lot of my aunts, uncles, and grandparents are old (thus highly susceptible to COVID). It was pretty stressful during Fall Quarter because one of my aunts got COVID and had to go the hospital for awhile"*

The *Location*, *Political*, and *COVID-19* codes represent categories of challenges that were persistent undercurrents in students' environments, causing worry and stress. These broader contexts in which students learned and lived certainly seemed to impact students' ability to engage with their classes and complete their coursework, though these were challenges not directly related to the course itself.

5.1.5 Well-being included physical and mental well-being. The fifth higher-level category of challenges students reported was that of personal well-being getting in the way of their learning, as is the case with the *Physical health*, *Mental health*, and *Isolation* codes. Students who mentioned *Physical health* sometimes described bouts of illness that caused them to fall behind in their courses or not feel well enough to do coursework.

> S-A-92: *"I had the worst case of mono for about the first three weeks of class, and now I am trying to catch up and relearn the basics of those first three weeks. I am having a difficult quarter"*

Other *Physical health* codes involved complications related to remote learning, such as sleep schedules being disrupted by having to attend synchronous classes in the middle of the night, or having physical symptoms from virtual learning through screens.

Students who reported *Mental health* challenges often wrote about stress, anxiety, and depression.

> S-B-41: *"Depression and anxiety, the pressure everyday"*

These kinds of challenges often co-occurred with reports of heavy course loads, difficult course content, or low engagement with others due to remote learning settings.

Somewhat similarly, *Isolation* codes had to do with students' socio-emotional well-being, and especially a lack of meaningful interactions with others.

> S-A-35: *"Being alone and lonely doing CS"*

Challenges that contained *Isolation* codes were often reported in conjunction with statements about the COVID-19 pandemic (due to quarantines, lockdowns, etc.) and remote learning.

> S-D-37: *"Being online is very isolating and does not allow for as much connection between students and instructors."*

All three these types of challenges – *Physical health*, *Mental health*, and *Isolation* – have to do with students' inner well-being, and health is an important prerequisite for effective learning.

5.1.6 Self-regulation involved motivation and time management. Finally, a sixth category of challenges students reported had to do with their own current self-regulation capacities and abilities. These were represented by the *Motivation* and *Time management* codes.

When students wrote about challenges that we coded as *Motivation*, they mentioned struggles with procrastination, distractions, focus, and feeling capable of completing coursework to their own standards.

> S-C-205: *"I have trouble finding the motivation to do school work nowadays."*

Oftentimes, *Motivation* codes co-occurred with *Remote learning* or *Isolation* codes, when students pointed out that their current environments were making it more difficult for them to focus or feel engaged within the course, or

students also wrote about the challenge of *Time management*, having to balance many demands and time constraints from their various roles as students, workers, family members, and humans in general. As a result, *Time management* codes co-occurred with many other codes, since students would often identify time as the challenge, then go on to describe the different facets of their lives that made time management difficult.

> S-C-16: *"I work about 30 hours a week as well as taking 17 credits this quarter, so time will be a challenge."*

Balancing all these demands could be difficult, especially when students struggled to set their own routines or to maintain a semblance of work/life balance.

> S-A-60: *"I only have so much time in the day/week for this class, other classes, and personal projects. It's honestly really rough to balance productivity and sanity :/"*

Both *Motivation* and *Time management* challenges had to do with students' current self-regulation capacities. Taken in context with many students reporting burnout from sources like the ongoing pandemic, competitive academic environments, and other stressors outside the class, it is perhaps not entirely surprising that students reported these kinds of challenges getting in the way of their learning.

5.1.7 Overall insights: Course-related feedback, external responsibilities, internal well-being. Overall, the six higher level categories of challenges students reported cover a wide range of potential learning difficulties. Some had to do with the course directly, as seen in the first section on course-related feedback codes. This set of difficulties is fairly similar to the feedback instructors might received from traditional methods (such as those described in the introduction), and perhaps the most "traditionally" actionable kinds of barriers to learning for instructors.

However, the other types of challenges reported through StudentAmp—challenges having to do with students' academic lives outside the course, their non-academic roles and broader contexts, their inner well-being and self-regulation skills—are particularly interesting to see. These latter categories of challenges might not show up through traditional student feedback methods. Instructors might never become aware of them if they exclusively used those kinds of methods, meaning that they likely would not be able to address them directly.

5.2 RQ2: Students' perceptions of sharing contextualized feedback

To better understand students' perceptions of sharing contextualized feedback, we conducted a thematic analysis on the transcripts of the interviews we conducted with students of minoritized groups (interviews previously described in Section 4.2.2). In the remainder of this section, we report four major themes students shared related to their perceptions of StudentAmp's purpose, as well as the ways in which different aspects of StudentAmp's design may have influenced what they shared or how they interacted with the tool.

5.2.1 Feedback was deemed important, even though purpose of StudentAmp was unclear. While interviewed students found feedback to be important to share, they expressed uncertainty about the purpose of StudentAmp. When we asked students what they thought the tool's intended purpose was, all five interviewed students expressed some uncertainty with two framing StudentAmp as a tool to improve the course. Students also compared StudentAmp to other feedback tools, such as direct emails, mid-term feedback, and end-of-term feedback. One student felt StudentAmp focused on and enabled conversation about a different context than other feedback they may have given:

> S-D-57: *"The feedback I said on StudentAmp was more 'What's going on with you? What are your challenges?' Whereas the feedback I gave for my instructor and my TA was 'Is the way they're teaching us helping? Are they getting back to us in time with questions?' I think they were just two different types of context in terms of what was being asked from us as students."*

Two participants noted how StudentAmp was unidirectional: Professors could get feedback from students, but not the other way around. One participant didn't expect much benefit from StudentAmp because of this lack of bidirectional feedback:

> S-C-88: *"[StudentAmp] is a place where professors could hear the voices of students, to some extent. But students could not hear what professor thought [...] in this case, it's not [a] participatory process, not exactly like that."*

5.2.2 Challenges beyond the scope of the class were worth sharing, but privacy mattered. When discussing value and risks of sharing challenges with StudentAmp, interviewed students talked about how differing goals and relationships with instructors affected what challenges they shared.

StudentAmp asked students to share the biggest challenge in their life, where that challenge could go beyond the scope of the class itself. Three interviewed students noted how their biggest challenge was often beyond the scope of the course, including S-D-57, who provided a metaphor of school as one of many "bubbles" in life:

> S-D-57: *"I've always thought that it's important that teachers or professors or people you interact with know a little bit about who you are and a little bit about what's in your surrounding bubbles. School is only one bubble of a student's life, so knowing all knowing a little bit about those other aspects about student life, you know family, emotional, work, relationships, friendships. Just knowing a little bit about those things can give you general knowledge of how it could be impacting the school bubble."*

Another interviewed student also recognized that life outside of school affects class experiences, but they had concerns with that overlap occurring:

> S-A-148: *"Stuff in your personal life definitely affects class, but because this is a school thing it makes me not want to 'cross those wires' almost. I'm worried about inappropriate timing, or something. I don't want to be 'that person.' Which is weird or doesn't really make sense, because that's what the survey is asking about. But I don't know. That can kind of be the 'little fear' in the back of your head."*

Students perceived multiple risks to privacy and safety that sometimes limited what they shared through StudentAmp. One student noted that they didn't want a person they lived with potentially seeing what they wrote while they interacted with the tool, and therefore opted not to share particular challenges. Another student noted how professors may not be the best person to respond to certain types of challenges, such as mental health:

> S-A-128: *"Maybe I could share these mental health [challenges] but, I don't know. I think I wouldn't just share those with a teacher, because you could share them with a professional who would be able to better help."*

Sharing about themselves made students feel vulnerable even when supposedly anonymous. One student justified this risk by feeling it was necessary as part of validating their challenges to the teaching team:

> S-A-148: *"I just hope that a professor would never think that I'm trying to take advantage of their kindness. [...] I wanted to let [my professor] know that I was serious, while staying anonymous, while asking for an extension specific for me. So I don't really know how you can satisfy all that."*

5.2.3 Demographic information was seen as an asset, although risk of re-identification existed. Interviewed students identified ways that sharing demographic information could help instructors understand the positionality of students facing different challenges. However, they also identified potential risks related to acceptance of minoritized identities and potential re-identification.

The students we interviewed generally felt that demographics could help instructors interpret challenges that students shared, with a few caveats. For instance, one student felt that instructors should have the cultural competence ([86]) to understand how societal structures affect learning experiences of students from minoritized groups:

> S-D-57: *" So if an as an instructor sees that the students who are BIPOC [...] and they're not doing as well as white students. I feel like a good informed instructor would know the racial understandings and the gender understandings as why certain groups with demographics will not be doing as great as other [groups]. Simply because of the world we live in, and the kind of.. structure our society is built upon. So I think a good instructor would know how to interpret that information and how to better help those students because they're all just trying to be at the same end goal."*

Another student found it relevant that StudentAmp asked demographic questions relating to other life commitments. Multiple interviewed students were searching for jobs while also taking this course, with one saying how StudentAmp helped connect job searching to course experience in a way that instructors previously had not:

> S-C-88: "*I appreciate that [StudentAmp] cares about whether we have jobs. Because I previously chatted with some other instructors and they said 'for job searches, that's kind of something different. I first care more about whether you learn well in this class.' Which makes me feel like I need to separate the job searching and course work. But they are not separate things. They're definitely things happening at the same time in my life.*"

Interviewed students also shared concerns related to their identities being seen as valid by teaching team members. One student shared uncertainties about how accepting the teaching team would be with regards to their minoritized gender identity:

> S-A-148: "*A risk might be [professors and TAs] don't take me seriously if they disagree with my identity or don't think my identity is valid [...] If this is a class I'm trying to do well [in] and take seriously, especially a class that's relevant to my major where I might see this professor again, or it matters a lot that I do well in this class, I'd be worried about not being taken seriously.*"

Another student identified the potential risk of re-identification. Even though courses were remote, teaching teams had additional information about students through their interactions with them as well as learning management systems. This information included a list of full names of all enrolled students. One student saw a potential risk of re-identification by connecting multiple pieces of demographic information with popularity of names in different cultures:

> S-A-128: "*With more [demographic information], like first-gen BIPOC, I feel like it would really narrow it down to a select few people [...] and there's tons of people who have similar names from certain regions. Like sometimes I can figure out where someone's from based off of their name.*"

5.2.4 Seeing others' challenges fostered community, but students questioned disrupt scores. Several interviewed students noted how seeing other students' feedback helped them feel less alone. Recall that students saw random pairs of their classmates challenges through StudentAmp (as described in Section 3.3.1). One student found that seeing challenges similar to their own made them feel less alone. Another student found the variety of challenges their classmates reported reassuring to see, especially during remote learning:

> S-A-148: "*It was nice to see that there's a variety [of challenges], that people are going through different things, or getting different things out of the class. But then when it's the same challenges as me, that's also reassuring, because then I am like 'okay I'm not the only one that's facing this right now, or having difficulty with this part of the class.'* "

Challenges that our interviewed students reported tended to have negative disrupt scores in the instructor view, suggesting that classmates found their challenges less disruptive compared to other challenges. StudentAmp aggregated meta-feedback responses into disrupt scores. Disrupt scores were the net number of times a classmate selected a given challenge over another challenge. The five interviewed students shared 13 challenges which had a median and mode disrupt score of -2, with the minimum being -6 and maximum being +2.

Students tended to question the aggregate disrupt score associated with challenges they reported, especially given how low they were. This did not bother some students, as they still personally felt their challenges were valid. However, other students recognized the impact that negative disrupt scores might have on instructors' awareness of needs of minoritized groups:

> S-C-88: "*As a user, when I see minus score, I would feel negative feelings definitely there was some judging behind it. And I understand probably people want to use this way to*

> *sort the results to help people browse information efficiently. But minority, disadvantaged, underrepresented people, they don't have many members or great numbers in the whole community. But still, they need to have their voice. It's not necessary because they are minority people and they have emergent needs, so other people would [...] probably be experiencing different things so that's my concern."*

One possible explanation for the observed variety in disrupt scores is that students interpreted the meta-feedback prompt differently than intended. To better understand how students perceived the request being made of them on the meta-feedback page, we asked interviewees to recall and reflect on their perceptions of it. During our first interviews, one student recalled interpreting the prompt as we intended (i.e., that they should choose which challenge would be more disruptive if they personally had it), two could not recall what they thought of the prompt, and the remaining two noted being surprised or confused by the prompt. Whereas the goal was to have students select the challenge they found more disruptive, one student interpreted the prompt as asking which challenge they also had and another as which challenge best represented the challenge they wrote.

Another potential explanation for low disrupt scores was that how students articulated their challenges affected how their classmates perceived them. One student thought that some students did not select their challenge because they used informal language (e.g., "whack" to describe a level of difficulty) and was not as verbose as some other students had been.

5.3 RQ3: Teaching teams used demographics to support perspective taking about challenges beyond the scope of the course

To analyze the interviews we conducted with teaching teams (previously described in Section 4.2.3), we conducted a collaborative thematic analysis on the transcripts with a sensitizing concepts of *teaching team interactions with StudentAmp* and *teaching team perceptions of student feedback*. Our approach was guided by the frequency of the topics raised by the teaching teams of the five courses as well as their relevance to answering our research question. In the remainder of this section, we report on the four major themes that arose from the interviews.

5.3.1 How teaching teams organized challenges reported in StudentAmp. To understand how teaching teams interpreted challenges that students reported in StudentAmp, we analyzed teaching teams' processes for creating and assigning labels to challenges. Professors and TAs could create custom labels to represent categories or groupings of challenges, and then assign them to challenges that they felt fit into those categories.

Some teaching teams focused on challenges most proximal to the course, such as those that dealt with course structure and course content, because they felt those challenges were most actionable. A TA in course A (T-A-6) read through all 139 responses in course A's week 4 feedback and labeled seven as "feature request,"[4] which were challenges they felt were actionable.

> T-A-6: *"'feature request' is a [label] name. It's just kind of actionable feedback that might help us make the course slightly easier for everybody and so it's more focused on [Course A] directly and things we do that may negatively impact how people learn."*

Rather than focusing on challenges that related to the course, other teaching teams looked at challenges more holistically. P-D worked with someone with qualitative research experience (outside of the research team) to analyze the first feedback session and developed the following labels and descriptions:

[4]For context, of 7 challenges that TA labeled as "feature request" we coded four as *Course structure*, with the others being coded as *Remote learning*, *Motivation*, and *Other* in our analysis of student-reported challenges (RQ1, see Table 2)

(1) *Mental Health*: Distinct from Stress below, a condition (such as Depression) that is experienced by the student
(2) *Stress + Time*: A (temporary) feeling resulting from excessive demands on time and energy, lack of time to complete work, challenges of work life balance
(3) *Motivation*: difficulties with being proactive, staying motivated, struggling to keep up, resisting burnout
(4) *Learning Environment*: issues with the physical space (e.g., noise, distractions), issues with internet connectivity, also isolation: feeling alone, lacking a sense of community
(5) *Group work*: difficulty collaborating with other students
(6) *New to Coding*: expression of intimidation, frustration, difficulty getting started, feeling "lost"

While P-D took a more holistic, top-down approach, a TA in course C (T-C-2) took a more bottom-up approach by looking at the first three pages of responses (75) in the first feedback session and creating labels that reflected causes of challenges. After TAs in course C worked together to label all 222 responses in the first feedback session, the three most commonly used labels were *online setting* (43), *workload* (35), and *social & collaboration* (32).

While some teaching teams focused on challenges that were more directly related to the course and more actionable (e.g. course structure), other teaching teams considered challenges even if they were beyond the control of the teaching team (e.g. lack of social interaction in remote learning).

5.3.2 How teaching teams considered demographics. In general, teachings teams tended not to consider demographics unless we prompted them during the interviews or unless they had prior training related to cultural competence. For instance, a TA in Course D (T-D-5) was familiar with speaking about equity and privilege from coursework in public health, and T-C-2 and T-C-3 were both currently enrolled in a course on educational equity and diversity. However, when members of the teaching team did consider demographics, they connected challenges to rich personas of students that deviated from expectations of dominant groups.

When Course B was reviewing challenges, they focused for several minutes on a specific challenge from their data: *"I'm unsure of my ability to train my brain to think this way"* (the student reporting this identified as part of several minoritized groups, including having mental and physical disabilities, taking their first programming course, and being a transfer student; disrupt score -2 = 3-5). When considering the challenge, a TA (T-B-1) who had the same gender identity as the student who wrote the challenge connected the challenge to their own experiences as a student and thought to remind students in lab section that "it's normal to struggle a little bit; it is challenging material and you're learning really fast." When prompted about the students' demographics, P-B focused on the transfer student label to identify implicit assumptions in the course design:

> P-B: *"a transfer student that makes me think of someone who's more likely than not probably coming in from a community college so may have the academic background, but doesn't necessarily know the way to navigate a four year institution effectively [...] physical disability minor [...] that could be someone who may be wearing a cast [...] And then severe mental disability could be any number of things as well, but definitely that would be something that would interfere with student's schedule or their ability to focus or their self esteem and their confidence and actually passing the course and and completing the assignments."*

P-B then went on to propose improvements such as clarifying how to use university email and access the course's learning management system, reassuring students that they could succeed in the course, and granting individual extensions on assignments.

Showing demographics did not necessarily translate to understanding on what to do with the information. When considering demographics for a challenge we labeled as job and course structure (*"I work 40 hours a week 8[A.M.] - 6[P.M.] so it can make it challenging to connect with TAs who only offer office hours during the middle of the day during the week..."*), P-E was uncertain how to consider the demographic information this student who identified as part of several minoritized groups, including being a transfer and first-generation student, working full-time, and job searching.

> P-E: *"In my head mentally, I still often see students as the standard undergrad 18 to 20 year old [...] living on campus or an apartment somewhere. The 'non-traditional' students as they're often frame are a different kind of aspect. I'm not sure what to do here now...in my head, this a challenge for everybody."*

P-E framed problems of students from minoritized groups as similar to students from dominant groups, but more severe. And while demographic information challenged the archetype of the "standard undergrad," P-E was unclear how to use this new information.

P-D and P-E reviewed feedback together. While P-E was unsure how to consider demographic information, P-D connected this challenge from a student from minoritized groups to systemic challenges at the university:

> P-D: *" what it feels like to read something like this is it is somewhere between heartbreaking and frustrating and angering. That is instructors were put in this really awful position where the university pressures people to take more courses than they can handle because [tuition] is so expensive."*

5.3.3 How teaching staff considered disrupt score. Disrupt scores were not taken literally, as the affordances of the interface design resulted in confusion amongst teaching teams and comparisons of such diverse challenges potentially confounding the aggregate disrupt score.

Because StudentAmp ordered challenges by disrupt score, affecting how teaching teams viewed and interacted with the data. StudentAmp ranked in each feedback session by disrupt score, resulting in teaching teams seeing challenges ordered from highest disrupt score to lowest. Disrupt score was shown as a number that is the difference between a positive number next to a thumbs up icon and negative number next to a thumbs down icon, as shown in Fig. 2e. It represented the net number of times a students selected that challenge over a random other one when asked to determine which challenge they found more disruptive.

One TA looked at the first three of nine pages (75 responses) and created and added labels to them, using the disrupt score as a stopping criteria:

> T-C-2: *"I just [went] over all the responses that are first three pages of responses and try to categorize and that basically settled all the tags [...] I think there is like a thumbs up, thumbs down. So I guess students get to like and dislike, or agree or disagree with certain statements. So up to page three, getting to a point [the disrupt score] is up one. So I think that's probably enough for telling what the students find most challenging."*

But later on, as they submitted a response as a student to explore the student view, the instructor questioned their interpretation of disrupt score as a measure of how many students related to a challenge:

> T-C-2: *"the net disruptive score [...] I don't like the minus sign. The first time I read this, I thought it means 'people do not agree with this.' [...] And then later on, when I actually [submitted] a student response, I pretended [I] was the student. I review the process, and then I realized it's asking which one is more disruptive instead of which one resonates more with your circumstance. I felt that there's a difference there and it's not really clear when I first viewed it."*

The thumbs up and down icons were similar with iconography used in many software interfaces to indicate rankings, but indicated something slightly different because selecting a response over another is not exactly the same as "upvoting," and choosing not to select a response is not exactly the same as "downvoting."

The professor for course A felt that the disrupt score was difficult to interpret in part because the diverse content of challenges made some comparisons uninformative. They gave an example of the disruptiveness of the global pandemic as being far greater than anything related to this class:

> P-A: *"The pandemic's huge, and to say something is less disruptive than a global pandemic that's not a very high bar. EVERYTHING should be less disruptive than a global pandemic. But it is nice that nobody is saying 'yeah there's something so wrong with this class, that is the biggest problem, even though there is a pandemic,' I suppose that's a win."*

Teaching teams really began to question low disrupt scores which corresponded to challenges they thought were disruptive. Of the responses which reported a challenge, responses with the lowest disrupt scores in each feedback session varied in content, but tended to be either vague or involved challenges that were not relatable to most students. The challenge with the lowest disrupt score in our dataset (-19=1-20) related to a phone being a distraction:

> S-D-69: *"My phone is the biggest challenge I am very addicted and it takes all of my focus during class times."*

TAs from course D labeled this challenge as *Motivation*. When reviewing this challenge, P-D wondered if other students were either downplaying the severity of the challenge or questioning its authenticity:

> P-D: *"Maybe people sort of downplaying the severity of that [challenge] or maybe that addiction is maybe something some people don't think is real."*

After hearing P-D's comment, P-E proposed an explanation wondering how the "non-clinical" language and focus on the phone may have caused peers to not take this challenge seriously:

> P-E: *"I wonder how many people are like 'oh haha yeah my phone is a joke too oh yeah no totally animal crossing is like definitely like my biggest distraction at the moment.' Whereas it's supposed to being like 'I have severe problems focusing on anything and I'm constantly spending time doom scrolling,' like there's actual things are going on there, but because of how [challenges] are framed and presented, they get read in very different ways."*

The disrupt score deviating from expectations caused P-D and P-E to propose alternative explanations for disrupt scores that deviated from their expectations. These explanations included students misunderstanding the meta-feedback prompts or not taking challenges seriously because of the way students reported challenges or because peers were unfamiliar with the challenges

P-A also noted a similar case where a challenge may have gotten a low disrupt score because it only affected a subset of students. For week 7 feedback in course A, we coded the challenge with the lowest disrupt score (-15 = 1-16) as location:

> S-A-117: *"Time difference"*

After seeing multiple challenges mentioning time zone differences with negative disrupt score, P-A acknowledged the impact of this challenge on a select few students:

> P-A: *"There's a few [responses] on timezone differences, but they are pretty consistently downvoted. Which I've heard enough now to believe that the timezone difference is a big deal for a small population of students."*

This theme of a low disrupt score for mental health related challenges also appeared in Course D. In week 1, a student who reported being part of several minoritized groups, including having a mental disorder which severely impacted their learning experience, stated a challenge related to severe depression and suicidal thoughts as part of week 1 feedback for Course D. This challenge received a disrupt score of -2 (6 - 8). In response, P-D talked about mental health in their next lecture and provided links to university resources to support students' mental health.

> P-D: *"So the students ranked severe depression and suicidal thoughts they rank that lower than the other thing you know eight out of 14 times which either means that students misunderstood the prompt or they misunderstand severe depression and suicidal thoughts."*

For week 7 feedback in Course A, S-A-202, who reported belonging to multiple minoritized groups including having a minor mental or social disorder which hindered their learning experience, stated a challenge related to depression and having to go to regular doctors appointments to manage their depression. This challenge had a disrupt score of 1, with 9 students selecting it over another random challenge and 8 students selecting a random challenge over this one. Upon seeing the lower disrupt score, P-A questioned the disrupt score and gave an explanation related to a lack of familiarity with depression:

> P-A: *"'My depression,' that's unfortunate. That should have a much disruption rating...I would bet you that the eight people who downvoted that don't have any history of mental illness or depression in themselves or their families, because if you know what that's like– that should be much higher."*

5.3.4 Teaching teams used StudentAmp to adjust course, training TAs, discuss systemic issues. While this evaluation focused on how students and teachers interpreted contextualized student feedback, we also identified three ways that teaching teams used this information.

First, teaching teams considered changes to course structure to be more accommodating to diverse students and their needs. Examples of this include supporting more community building amongst students who felt isolated by remote learning, making deadlines and office hours more flexible to accommodate students from different time zones and those who worked jobs, and supporting students' mental health by raising awareness of free university resources and finding new ways to express empathy.

Another way teaching teams used StudentAmp was for development opportunities for the teaching teams themselves. While we framed StudentAmp as a way to improve teaching practices by responding to student challenges, P-C saw StudentAmp more of as an opportunity to discuss diverse student experiences with course TAs to build empathy.

> P-C: *"The point of a StudentAmp survey is not to collect data on students or even to improve instruction (as in formative course evaluations), but rather to amplify student experiences that might otherwise fall between the cracks. In response, the instructor's 'fireside chat' offers a natural mechanism for the instructor to recognize and validate student experiences revealed through StudentAmp."*

Finally, StudentAmp did foster some discussion about systemic issues which extended beyond the course and even beyond the university. Teaching teams felt limited in what actions they could take in the middle of their large remote courses, but they still used StudentAmp to discuss broader systemic challenges that their students faced:

> P-D: *"There are certain dials that [professors] can turn and they're still contextualized within the university system where all the other courses they're taking have firm deliverables and it's contextualized within a broader social and economic system in which, if they don't get a good job they can't go to the doctor later or pay off their huge loans"*

6 DISCUSSION & CONCLUDING REMARKS

In this paper, we designed, developed, and evaluated StudentAmp, a student feedback tool that supported equity-oriented goals by asking students to report challenges that may be beyond the immediate scope of the course and contextualizing those challenges with self-reported demographic information as well as an aggregate score reflecting peer perceptions of challenges. We evaluated StudentAmp with five large computing courses (150 - 750 students) that were taught remotely during dual pandemics of COVID-19 and racial injustice. We found that students used StudentAmp to share challenges beyond the scope of the course, including challenges in non-academic roles and challenges related to the well-being of themselves, their families, and their peers. Interviewed students identified a tension between wanting to share more information about their lives to justify their needs while also wanting to preserve their anonymity and safety. Teaching teams used this contextualized feedback to consider not just challenges, but also the positionality of the student reporting the challenges. Taken together, this paper contributes a design exploration into how contextualizing student feedback can support equity-oriented goals in large, remote computing courses.

In this section, we describe multiple ways to interpret our findings. We focus in particular on the primary tension that this study: How to support equity-oriented goals by contextualizing student feedback while also ensuring the privacy, well-being, and trust of students, especially students of minoritized groups.

6.1 Limitations: Self-selection bias

One interpretation of our findings is that they may lack validity because of self-selection bias throughout the study. Indeed, participants self-selected into the study at multiple phases, with instructors choosing to participate in this study and use StudentAmp, a subset of students choosing to share feedback with StudentAmp, unequal usage of Student across the five courses in the study, and five students choosing to interview with us. This type of bias was likely exacerbated by the context during which we conducted this research, in which students were still adjusting to remote learning and trying to do so during dual pandemics of COVID-19 and racial injustice. It was partially due to these challenging times that we decided to conduct this research, because students were facing new or worsening challenges to learning and because teaching teams needed to understand how to support them, but lacked the capacity to do so.

We tried to mitigate self-selection bias by compensating instructors, TAs, and students for their time and allowing for flexibility in regards to what participants were comfortable disclosing and with scheduling. We also conducted repeated (at least two) interviews with each teaching team and student interview participant. A follow-up survey with all students could have provided corroboration to themes that we identified in our interviews, and future work on this topic would do well to explore more deeply the saturation and relative frequency of the themes we surfaced. As a result, while self-selection bias is indeed a limitation of this study, we can still nevertheless interpret our findings as a contribution to a larger body of knowledge that seeks to understand how to design contextualized student feedback for equity-oriented goals.

6.2 StudentAmp focuses on students' experiences in a broader context

Another interpretation is that StudentAmp is similar to other feedback tools that already exist. All five courses that used StudentAmp also used other common student feedback techniques, such as online surveys, direct conversations, mid-term feedback, and end-of-term feedback. However, interviewed students felt that StudentAmp was different than these techniques because it afforded them an opportunity to share feedback about their experiences beyond the immediate scope of the course, which most other feedback tools focused on. The anonymity of StudentAmp enabled them to be more open and vulnerable in ways that identifiable techniques such as direct conversations and emails do not afford. We found that seeing these broader challenges as well as demographic information helped teaching teams discuss student challenges that went beyond course structure and affected certain groups of students disproportionately (e.g. mental well-being, timezone differences, and job or familial commitments). In these cases, StudentAmp enabled broader and more contextualized feedback, and future work can explore how to better integrate aspects of StudentAmp into different pedagogical practices with student feedback.

6.3 StudentAmp avoids reducing students to labels, supports perspective Taking

Another interpretation is that StudentAmp is harmful through its reductions of people's diverse lived experiences. Data technologies can promote material, symbolic, and other violences by reducing people to broad demographic groups and promoting incremental changes that do not address larger systemic issues [39, 40]. To avoid data violence related to stereotyping, we designed StudentAmp to show multiple dimensions of demographic information, enabling consideration of intersectional identities for perspective taking. StudentAmp enabled aggregation and reduction of challenges (through user-defined labels), but not by demographic groups. These design decisions required teaching teams to consider challenges as contextualized by multiple dimensions of student demographics. We found that discussing demographic information was challenging for teaching teams, with members with more cultural competence (e.g. training in culturally responsive teaching) more able to lead these discussions.

Showing high-dimensional demographic information likely enabled consideration of intersectional identities in perspective taking. Rather than see one or two demographic labels (e.g. gender and ethnicity), teaching teams could see up to nine (enumerated in section 3.3.1). Making visible multiple demographic features at once provided a mechanism to discourage simple and harmful stereotyping and instead supported more nuanced perspective taking. Furthermore, this feature provided teaching teams the option to talk about aspects of demographics they were most comfortable with. For example, we noticed in multiple interviews how teaching teams discussed that a student was a transfer or first-generation and how this was more common amongst BIPOC students. All five professors and many TAs identified as coming from dominant groups, so speaking about some aspects of demographics may have been less comfortable with (e.g. gender, ethnicity) and other aspects of demographics may have been more comfortable with (e.g. transfer student, first-generation student). From our data, it is unclear what the impact of using more comfortable demographic labels in discussions about equity might be, since it could be an inroad to starting "more difficult" discussions, or it could obscure more direct systemic issues about gender and ethnicity. This could be an exploration for future work.

6.4 StudentAmp affords some anonymity, but privacy risks still exist

Another interpretation is that StudentAmp poses a privacy risk to students. In particular, multiple demographic labels make common privacy guarantees such as k-anonymity [81] impossible. Furthermore, background knowledge attacks [52], where an adversary (e.g. member of teaching team) uses information from other data sources to re-identify a respondent may be of particular risk within a learning context. Given how instructors, TAs, and students know and interact with each other frequently and instructors and TAs have additional information about students (e.g. names and pictures). We mitigated this risk by recruiting culturally competent instructors and not enabling aggregation or filtering by demographic group in the StudentAmp interface. However, one interviewed student noted how re-identification could still be possible with background knowledge (e.g. name to infer ethnicity or gender) and how this could be especially dangerous if an instructor or TA was less culturally competent. Future work can explore how to reduce the risk of re-identification by disclosing demographic information in a non-uniform way that provides relevant context for specific challenges (e.g. mention mental disabilities when a challenge relates to mental health, but not familial language) or adjusting their buckets (e.g. use BIPOC label if there are only a few Black/African students but more Hispanic/Latinx students). McDonald's framing of privacy from a vulnerability perspective could guide improvements to privacy and safety of minoritized groups in particular [56].

6.5 Organizing information in a scalable, equitable, and privacy-preserving way is an open design space

Yet another interpretation is that StudentAmp was not equitable because it did not draw attention to the needs of minoritized groups. We tried to show the information about challenges in an equitable way by doing two things: First, StudentAmp attempted to call attention to minoritized groups' needs by showing demographic information with challenges only if that person was from a minoritized group. This enabled perspective-taking behavior, as discussed in Section 6.3. Second, StudentAmp also ranked challenges by disrupt score in an attempt to organize challenges by disruptiveness, but teaching teams found disrupt scores to be a confounded measure.

The goal of the disrupt score was to introduce a mechanism that organized challenges by disruptiveness and not frequency or commonness, as minoritized groups could make up a small proportion of students and/or have unique challenges. In our interviews with teaching teams, we found multiple instances of them being uncertain of how to interpret the apparent contradiction of low disrupt scores for seemingly severe challenges, such as mental health concerns. Potential explanations include StudentAmp providing an unclear prompt or explanation for the meta-feedback task, as evidenced by multiple interviewed students interpreting meta-feedback prompts in different ways and being uncertain about what they were doing. Asking a student to consider the disruptiveness of challenges that they may never have encountered for the purposes of calculating the disrupt score is also a difficult task. Future work can explore potential improvements including the effect of different challenge selection procedures (e.g. asking students to consider challenges that are similar to their own, or from people from similar demographic groups) or contextualizing meta-feedback with demographic information or other information to enable more informed consideration.

An improved disrupt score is only one of many potential mechanisms to support the organization of information in an equitable yet efficient way while also ensuring privacy and well-being of minoritized groups especially. A large remote course with hundreds of students is a dynamic environment where students' needs must be met in a timely manner. Efficiency in data collection and analysis is key to taking action. Furthermore, students of minoritized groups may face challenges that are unique and unlike challenges their peers have, so there must be a way to organize information so perspectives of minoritized groups are not lost. Lastly, addressing students' needs often involves vulnerability and asking students to share information about their experiences both in and out of the class. Because students of minoritized groups are perhaps most vulnerable to privacy violations, we must ensure the privacy and well-being of minoritized groups in particular. These tensions present a rich design space for future work that values equity and human well-being while also wrangling with the pragmatics involved with the need to scale.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 12566082, 1762114, 1539179, 1703304, 1836813, 2031265, 2100296, 2122950, 2137834, 2137312, the Google Cloud Research Credits program, and unrestricted gifts from Microsoft, Adobe, and Google. This study was pre-approved with exempt status by the University of Washington Institutional

Review Board (IRB) as STUDY00008871. Supplemental material can be found at https://github.com/codeandcognition/archive-2022cscw-xie.

REFERENCES

[1] Farshid Anvari, Deborah Richards, Michael Hitchens, and Hien Minh Thi Tran. 2019. Teaching user centered conceptual design using cross-cultural personas and peer reviews for a large cohort of students. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET '19)*. IEEE Press, Montreal, Quebec, Canada, 62–73. https://doi.org/10.1109/ICSE-SEET.2019.00015

[2] Chris Argyris and Donald A Schön. 1978. *Organizational Learning: A Theory of Action Perspective*. Addison-Wesley Publishing Company.

[3] Chris Argyris and Donald A Schön. 1996. *Organizational Learning II: Theory, Method, and Practice*. Addison-Wesley Publishing Company.

[4] Cynthia L. Bennett and Daniela K. Rosner. 2019. The Promise of Empathy: Design, Disability, and Knowing the "Other". In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300528

[5] Aikaterini Bourazeri and Simone Stumpf. 2018. Co-designing smart home technology with people with dementia or Parkinson's disease. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction (NordiCHI '18)*. Association for Computing Machinery, New York, NY, USA, 609–621. https://doi.org/10.1145/3240167.3240197

[6] Glenn A Bowen. 2006. Grounded Theory and Sensitizing Concepts. *International Journal of Qualitative Methods* 5, 3 (Sept. 2006), 12–23. https://doi.org/10.1177/160940690600500304

[7] Shannon T Brady, Geoffrey L Cohen, Shoshana N Jarvis, and Gregory M Walton. 2020. A brief social-belonging intervention in college improves adult outcomes for black Americans. *Science advances* 6, 18 (May 2020), eaay3689. https://doi.org/10.1126/sciadv.aay3689

[8] Emeline Brulé and Katta Spiel. 2019. Negotiating Gender and Disability Identities in Participatory Design. In *Proceedings of the 9th International Conference on Communities & Technologies - Transforming Communities (C&T '19)*. Association for Computing Machinery, New York, NY, USA, 218–227. https://doi.org/10.1145/3328320.3328369

[9] Kenneth M Burke. 2010. Distributed leadership and shared governance in post-secondary education. *Management in Education* 24, 2 (April 2010), 51–54.

[10] Margaret Burnett, Simone Stumpf, Jamie Macbeth, Stephann Makri, Laura Beckwith, Irwin Kwan, Anicia Peters, and William Jernigan. 2016. GenderMag: A method for evaluating software's gender inclusiveness. *Interacting with Computers* 28, 6 (2016), 760–787. https://doi.org/10.1093/iwc/iwv046

[11] Center for Educational Leadership. 2014. *Creating a Theory of Action for Improving Teaching and Learning*. Technical Report. University of Washington.

[12] Sin Wang Chong. 2021. Reconsidering student feedback literacy from an ecological perspective. *Assessment & Evaluation in Higher Education* 46, 1 (Jan. 2021), 92–104. https://doi.org/10.1080/02602938.2020.1730765

[13] Derrick Coetzee, Armando Fox, Marti A Hearst, and Björn Hartmann. 2014. Should your MOOC forum use a reputation system?. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1176–1187. https://doi.org/10.1145/2531602.2531657

[14] Ira J Cohen. 1989. Structuration Theory and Social Praxis. In *Structuration Theory: Anthony Giddens and the Constitution of Social Life*, Ira J Cohen (Ed.). Macmillan Education UK, London, 9–55.

[15] James S Coleman. 1986. Social Theory, Social Research, and a Theory of Action. *Amer. J. Sociology* 91, 6 (1986), 1309–1335.

[16] Alison Cook-Sather. 2018. Listening to equity-seeking perspectives: how students' experiences of pedagogical partnership can inform wider discussions of student success. *Higher education research & development* 37, 5 (July 2018), 923–936. https://doi.org/10.1080/07294360.2018.1457629

[17] Alison Cook-Sather. 2020. Respecting voices: how the co-creation of teaching and learning can support academic staff, underrepresented students, and equitable practices. *Higher Education* 79, 5 (May 2020), 885–901. https://doi.org/10.1007/s10734-019-00445-w

[18] Arthur H Copeland. 1951. A reasonable social welfare function. Seminar on Applications of Mathematics to Social Sciences.

[19] Sasha Costanza-Chock. 2020. *Design Justice: Community-led practices to build the worlds we need*. MIT Press.

[20] National Research Council. 2012. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. The National Academies Press, Washington, DC.

[21] Kimberlé Williams Crenshaw. 1994. Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. In *The Public Nature of Private Violence*, Martha Albertson Fineman and Rixanne Mykitiuk (Eds.).

Routledge, 93–118.

[22] Mark H. Davis. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology* 44, 1 (1983), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

[23] Ruth Dunn. 2021. *Minority Studies.* LibreTexts.

[24] Barbara Ericson and Mark Guzdial. 2014. Measuring demographics and performance in computer science education at a nationwide scale using AP CS data. In *Proceedings of the 45th ACM technical symposium on Computer science education* (Atlanta, Georgia, USA) *(SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 217–222.

[25] Karl Anders Ericsson and Herbert Alexander Simon. 1993. *Protocol Analysis: Verbal Reports as Data Revised Edition.* The MIT Press.

[26] Y Fan, L J Shepherd, E Slavich, D Waters, M Stone, R Abel, and E L Johnston. 2019. Gender and cultural bias in student evaluations: Why representation matters. *PLoS One* 14, 2 (Feb. 2019), e0209749.

[27] Gail E FitzSimons. 2011. A Framework for Evaluating Quality and Equity in Post-Compulsory Mathematics Education. In *Mapping Equity and Quality in Mathematics Education*, Bill Atweh, Mellony Graven, Walter Secada, and Paola Valero (Eds.). Springer Netherlands, Dordrecht, 105–121.

[28] Floyd J Fowler, Jr. and Thomas W Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error.* SAGE.

[29] Mark Freeman, Paul Blayney, and Paul Ginns. 2006. Anonymity and in class learning: The case for electronic response systems. *Australasian Journal of Educational Technology* 22, 4 (Nov. 2006).

[30] Nichole M Garcia, Nancy López, and Verónica N Vélez. 2018. QuantCrit: rectifying quantitative methods through critical race theory. *Race Ethnicity and Education* 21, 2 (March 2018), 149–157.

[31] Emily Gould and Rachael Swell. 2011. Managing labels. https://docs.github.com/en/github/managing-your-work-on-github/managing-labels. Accessed: 2021-4-2.

[32] Colin Gray, Seda Yilmaz, Shanna Daly, Colleen Seifert, and Richard Gonzalez. 2015. Idea Generation Through Empathy: Reimagining the 'Cognitive Walkthrough'. *2015 ASEE Annual Conference and Exposition* (June 2015). https://lib.dr.iastate.edu/industrialdesign_conf/10

[33] Pamela Grimm. 2010. Social Desirability Bias. In *Wiley International Encyclopedia of Marketing*, Jagdish Sheth and Naresh Malhotra (Eds.). Vol. 50. John Wiley & Sons, Ltd, Chichester, UK, 537. https://doi.org/10.1002/9781444316568.wiem02057

[34] David Hammer and Leema K Berland. 2014. Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences* 23, 1 (2014), 37–46.

[35] Kay Hempsall. 2014. Developing leadership in higher education: perspectives from the USA, the UK and Australia. *Journal of Higher Education Policy and Management* 36, 4 (July 2014), 383–394.

[36] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466, 7302 (July 2010), 29.

[37] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behav. Brain Sci.* 33, 2-3 (June 2010), 61–83; discussion 83–135.

[38] Charles G Hill, Maren Haag, Alannah Oleson, Chris Mendez, Nicola Marsden, Anita Sarma, and Margaret Burnett. 2017. Gender-Inclusiveness Personas vs. Stereotyping: Can We Have it Both Ways?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6658–6671.

[39] Anna Lauren Hoffmann. 2018. Data Violence and How Bad Engineering Choices Can Damage Society. https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4. Accessed: 2021-7-02.

[40] Anna Lauren Hoffmann. 2020. Terms of inclusion: Data, discourse, violence. *New Media & Society* (Sept. 2020), 1461444820958725. https://doi.org/10.1177/1461444820958725

[41] Meredith I Honig, Michael A Copland, Lydia Rainey, Juli Anna Lorton, and Morena Newton. 2010. *Central Office Transformation for District-wide Teaching and Learning Improvement.* Technical Report. University of Washington.

[42] Aleata Hubbard Cheuoua. 2021. Confronting Inequities in Computer Science Education: A Case for Critical Theory. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Virtual Event, USA) *(SIGCSE '21)*. Association for Computing Machinery, New York, NY, USA, 425–430.

[43] Magdalena Jara and Harvey Mellar. 2010. Quality enhancement for e-learning courses: The role of student feedback. *Comput. Educ.* 54, 3 (April 2010), 709–714.

[44] Daniel B King and Shivani Joshi. 2008. Gender Differences in the Use and Effectiveness of Personal Response Devices. *J. Sci. Educ. Technol.* 17, 6 (Dec. 2008), 544–552.

[45] Michael S Knapp, Michael A Copland, Meredith I Honig, Margaret L Plecki, and Bradley S Portin. 2010. *Learning-Focused Leadership and Leadership Support: Meaning and Practice in Urban Systems.* Technical Report. University of Washington.

[46] Amy J Ko, Alannah Oleson, Neil Ryan, Yim Register, Benjamin Xie, Mina Tari, Matthew Davidson, Stefania Druga, and Dastyni Loksa. 2020. It is time for more critical CS education. *Commun. ACM* 63, 11 (Nov. 2020), 31–33.

[47] Mary Ellen Kondrat. 2002. Actor-Centered Social Work: Re-visioning "Person-in-Environment" through a Critical Theory Lens. *Soc. Work.* 47, 4 (Oct. 2002), 435–448.

[48] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S Bernstein, and Scott R Klemmer. 2015. Talkabout: Making Distance Matter with Small Groups in Massive Classes. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1116–1128. https://doi.org/10.1145/2675133.2675166

[49] Harrison Kwik, Benjamin Xie, and Amy J Ko. 2018. Experiences of Computer Science Transfer Students. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. ACM Press, 115–123.

[50] Colleen M Lewis, Niral Shah, and Katrina Falkner. 2019. Equity and diversity. In *The Cambridge Handbook of Computing Education Research*, Sally A Fincher and Anthony V Robins (Eds.). Cambridge University Press, 481–510.

[51] Sarah Lewthwaite and David Sloan. 2016. Exploring Pedagogical Culture for Accessibility Education in Computing Science. In *Proceedings of the 13th Web for All Conference (W4A '16)*. ACM, New York, NY, USA, 3:1–3:4. https://doi.org/10.1145/2899475.2899490

[52] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data* 1, 1 (March 2007), 3–es.

[53] Martin N Marger. 2015. *Race and Ethnic Relations: American and Global Perspectives, 10th Edition.* Cengage.

[54] Herbert W Marsh and Lawrence A Roche. 1997. Making Students' Evaluations of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias, and Utility. *American Psychologist Association* 52, 11 (Jan. 1997).

[55] Robert McCartney, Jonas Boustedt, Anna Eckerdal, Kate Sanders, and Carol Zander. 2017. Folk Pedagogy and the Geek Gene: Geekiness Quotient. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) *(SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 405–410.

[56] Nora McDonald, Karla Badillo-Urquiola, Morgan G Ames, Nicola Dell, Elizabeth Keneski, Manya Sleeper, and Pamela J Wisniewski. 2020. Privacy and Power: Acknowledging the Importance of Privacy Research and Design for Vulnerable Populations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8.

[57] Paola Medel and Vahab Pournaghshband. 2017. Eliminating Gender Bias in Computer Science Education Materials. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education - SIGCSE '17*. ACM Press, Seattle, Washington, USA, 411–416.

[58] Christopher Mendez. 2020. *The InclusiveMag Method: A Start Towards More Inclusive Software for Diverse Populations.* Ph.D. Dissertation. Oregon State University. https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/8910k155q Publisher: Oregon State University.

[59] Christopher Mendez, Lara Letaw, Margaret Burnett, Simone Stumpf, Anita Sarma, and Claudia Hilderbrand. 2019. From GenderMag to InclusiveMag: An Inclusive Design Meta-Method. In *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 97–106. https://doi.org/10.1109/VLHCC.2019.8818889

[60] Christof Nachtigall, Ulf Kröhne, Ulrike Enders, and Rolf Steyer. 2008. Causal effects and fair comparison: Considering the influence of context variables on student competencies. In *Assessment of Competencies in Educational Contexts*, Johannes Hartig, Eckhard Klieme, and Detlev Leutner (Eds.). Hogrefe Publishing, 315–336.

[61] National Academies of Sciences and Medicine. 2018. *Assessing and Responding to the Growth of Computer Science Undergraduate Enrollments.* The National Academies Press, Washington, DC.

[62] Timothy Neate, Aikaterini Bourazeri, Abi Roper, Simone Stumpf, and Stephanie Wilson. 2019. Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300880

[63] Alannah Oleson, Christopher Mendez, Zoe Steine-Hanson, Claudia Hilderbrand, Christopher Perdriau, Margaret Burnett, and Andrew J Ko. 2018. Pedagogical Content Knowledge for Teaching Inclusive Design. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. Association for Computing Machinery, New York, NY, USA, 69–77.

[64] Alannah Oleson, Meron Solomon, and Amy J. Ko. 2020. Computing Students' Learning Difficulties in HCI Education. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–14. https://doi.org/10.1145/3313831.3376149

[65] Elizabeth Patitsas, Jesse Berlin, Michelle Craig, and Steve Easterbrook. 2016. Evidence That Computer Science Grades Are Not Bimodal. In *Proceedings of the 2016 ACM Conference on International Computing Education Research (ICER '16)*. ACM, New York, NY, USA, 113–121.

[66] Michael Quinn Patton. 2014. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice.* SAGE Publications.
[67] Heather E Price. 2019. Large-Scale Datasets and Social Justice: Measuring Inequality in Opportunities to Learn. In *Research Methods for Social Justice and Equity in Education*, Kamden K Strunk and Leslie Ann Locke (Eds.). Springer International Publishing, Cham, 203–215.
[68] Cynthia Putnam, Maria Dahman, Emma Rose, Jinghui Cheng, and Glenn Bradford. 2015. Teaching Accessibility, Learning Empathy. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15.* ACM Press, Lisbon, Portugal, 333–334. https://doi.org/10.1145/2700648.2811365
[69] John T E Richardson. 2005. Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education* 30, 4 (Aug. 2005), 387–415. https://doi.org/10.1080/02602930500099193
[70] Kevin Robinson, Keyarash Jahanian, and Justin Reich. 2018. Using Online Practice Spaces to Investigate Challenges in Enacting Principles of Equitable Computer Science Teaching. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) *(SIGCSE '18).* Association for Computing Machinery, New York, NY, USA, 882–887.
[71] Kimberley Rogers, Isabel Dziobek, Jason Hassenstab, Oliver T. Wolf, and Antonio Convit. 2007. Who Cares? Revisiting Empathy in Asperger Syndrome. *Journal of Autism and Developmental Disorders* 37, 4 (April 2007), 709–715. https://doi.org/10.1007/s10803-006-0197-8
[72] Karen Rosenblum and Toni-Michelle Travis. 2015. *The Meaning of Difference: American Constructions of Race and Ethnicity, Sex and Gender, Social Class, Sexuality, and Disability.*
[73] Donald G Saari and Vincent R Merlin. 1996. The Copeland method. *Economic theory* 8, 1 (Feb. 1996), 51–76. https://doi.org/10.1007/bf01212012
[74] Kimberly A Scott, Kimberly M Sheridan, and Kevin Clark. 2015. Culturally responsive computing: a theory revisited. *Learn. Media Technol.* 40, 4 (Oct. 2015), 412–436.
[75] Kimberly A Scott and Mary Aleta White. 2013. COMPUGIRLS' Standpoint: Culturally Responsive Computing and Its Effect on Girls of Color. *Urban Education* 48, 5 (Sept. 2013), 657–681.
[76] Simone G. Shamay-Tsoory, Judith Aharon-Peretz, and Daniella Perry. 2009. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132, 3 (March 2009), 617–627. https://doi.org/10.1093/brain/awn279
[77] Katta Spiel, Oliver Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: A guide for HCI researchers. *ACM Interactions* 26, 4 (2019).
[78] Philip Stark, Richard Freishtat, and Carol Lauer. 2014. An Evaluation of Course Evaluations. *ScienceOpen Research* (Sept. 2014).
[79] Claude Steele. 2011. Stereotype Threat and African-American Student Achievement. In *The Inequality Reader* (2 ed.). Routledge, 276–281.
[80] Simone Stumpf, Anicia Peters, Shaowen Bardzell, Margaret Burnett, Daniela Busse, Jessica Cauchard, and Elizabeth Churchill. 2020. Gender-Inclusive HCI Research and Design: A Conceptual Review. *Foundations and Trends® in Human–Computer Interaction* 13, 1 (March 2020), 1–69. https://doi.org/10.1561/1100000056 Publisher: Now Publishers, Inc.
[81] Latanya Sweeney. 2002. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems* 10, 5 (2002).
[82] Burçin Tamer and Jane Stout. 2016. *Recruitment and Retention of Undergraduate Students in Computing: Patterns by Gender and Race/Ethnicity.* Technical Report. Computing Research Association.
[83] Phil Turner and Susan Turner. 2011. Is stereotyping inevitable when designing with personas? *Design Studies* 32, 1 (Jan. 2011), 30–44. https://doi.org/10.1016/j.destud.2010.06.002
[84] Gregory Walton and Geoffrey Cohen. 2007. A question of belonging: Race, social fit, and achievement. *Journal of personality and social psychology* 92 (02 2007), 82–96. https://doi.org/10.1037/0022-3514.92.1.82
[85] Jayce R Warner, Joshua Childs, Carol L Fletcher, Nicole D Martin, and Michelle Kennedy. 2021. Quantifying Disparities in Computing Education: Access, Participation, and Intersectionality. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Virtual Event, USA) *(SIGCSE '21).* Association for Computing Machinery, New York, NY, USA, 619–625.
[86] Alicia Nicki Washington. 2020. When Twice as Good Isn't Enough: The Case for Cultural Competence in Computing. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '20).* Association for Computing Machinery, New York, NY, USA, 213–219.
[87] Max Weber. 1948. *From Max Weber: Essays in Sociology.* Vol. 33. Routledge.
[88] S Christian Wheeler and Richard E Petty. 2001. The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychol. Bull.* 127, 6 (Nov. 2001), 797–826.

[89] Saijing Zheng, Mary Beth Rosson, Patrick C Shih, and John M Carroll. 2015. Understanding Student Motivation, Behaviors and Perceptions in MOOCs. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1882–1895. https://doi.org/10.1145/2675133.2675217
[90] Tukufu Zuberi. 2001. *Thicker Than Blood: How Racial Statistics Lie.* U of Minnesota Press.
[91] Stuart Zweben and Betsy Bizot. 2019. *Taulbee Survey.* Technical Report. Computing Research Association.

Received July 2021; revised November 2021; accepted November 2021