Abstract: Stakeholders' Interpretations of Data for Equitable Computing Education

Benjamin Xie
Ph.D. in Information Science

1. Introduction: Connecting data with domain expertise for equitable learning. Computing education has seen a boom in interest and enrollment that has resulted in the use of more scalable, data-driven technologies to support learning experiences (e.g. online learning systems, student feedback tools, and auto-graders). These data-driven tools often serve to scale learning experiences by standardizing them to serve the majority and/or relying on data from students of dominant identity groups (e.g. able white and Asian men in the US) to train adaptive learning tools. But not students are not all the same, and these "scalable" experiences will often fail to serve and even harm students in *minoritized* groups, groups that have been excluded or isolated because of societal structures (e.g. systemic racism, exclusions, oppression) [21, 22]. Minoritized groups in computing, as regretably consistent with much of academia, include African-American/Black, Hispanic/Latinx, Indigenous/Native American, Pacific Islander, women, non-binary, first-generation students, and transfer students [10].

Data can help inform and enable timely action by substantiating nuanced patterns related to disparities and bias so we can consider how to appropriately adjust resources [18]. While equity is a goal that is typically too complex and situated to easily measure with data [30], it cannot exist without first analyzing data to understand existing disparities and biases [18]. Data can measure the existence and extent of disparities and biases, but judgement is required to interpret and act on data-related findings [33]. So enabling stakeholders with domain expertise to interpret and use data can help them take more informed, timely, and equitable action.

In this dissertation, I explored how to use data to support equity by putting it in the hands of stakeholders situated in learning experiences (students, teachers, curriculum designers). I conducted three studies to explore how to design interactions that enable stakeholders to connect their domain expertise to data about learners' experiences. From these design explorations, I developed a unifying framework to inform how we design interactions with data to support equitable learning experiences.

2. Background: Using data for equity is a sociotechnical design problem. To understand how to support equity, we must first understand how inequities exist. To do this, I took a critical stance by stepping back to *problematize* dominant historical and cultural norms in computing education. This followed Michel Foucault's tradition of problematics [8], which has since been adopted into the method of *problematization* by James D. Marshall [12]. Problematization frames knowledge within malleable political, social, economic, historic and cultural norms that we should first investigate by "stepping back" [12, 8]. It also requires considering power relationships at a systemic level to explore how *dominant* systems came to be and how they operate to oppress *minoritized* groups. Examples of problems include the formation of a scientific domain, political structure, and moral practice. I used problematization as a framing to consider and critically question dominant systems, discourses, and actions to identify more equitable alternatives. By doing so, we could consider how to use data to inform stakeholders on who is or is not being served by existing systems to imagine more equitable alternatives.

Improving equity would require not just improving access to computing education, but also supporting successful participation and achievement by diverse students learning computing [11]. 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 [23]), affect students' sense of belonging and identity (e.g. instructional materials promoting gender bias [14]), 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) [11]. Inequities arise when structures and norms fail to include or serve students of minoritized groups.

Critical to the concept of equity is a consideration of historical injustices. This differentiates equity from related concepts of *equality* and *inclusion*. Understanding equality or lack thereof is often a necessary precursor to supporting equity [18]. This dissertation explores how to use data for equity-oriented purposes by having stakeholders interpret data to inform future actions.

Using student data comes with inherent risks that computing education research (CER) and learning analytics have previously identified. Risks of using data include violations of privacy and reduction of nuanced lived experiences to low dimensional data. Privacy involves consideration of trust, transparency, student agency/control over data, security, and accountability/reasonable care of data [17, 21]. A critical perspective on the use of data and learning analytics considers the role of power, impact of surveillance, and identity of students as transient, temporal, and context-bound constructs [25]. Addressing privacy concerns ultimately involves context-rich information that considers asymmetrical power relationships [19].

Another concern about the use of data is reduction of nuanced students and their identities to low dimensional data. Computing education and Learning Analytics researchers often adhere to a postivistic framing that assumes that "objective" aggregation and quantification of data can predict individual behavior to sufficient accuracy [28]. But this assumption that aggregate behavior reflects individuals is contrary to goals of equity because it purports that behaviors of minoritized groups reflect that of dominant groups. A more critical and equitable lens is that of *intersectionality*, denoting the various ways in which ethnicity and gender (and other demographic labels) interact with social structures to shape the peoples' lived experiences [4].

Traditional statistical methods when used within a positivist framing can perpetuate norms of dominance and oppression by treating data as complete, unbiased measurements that we can consider in isolation [6]. Statistical methods have the most "power" when we accept reductions of the world to dichotomies and accept a regression to a common average. But these false dichotomies typically do not reflect the diversity of people, and the assumption that dominant norm reflects minoritized groups are often what equity-oriented actions must overcome. Because stakeholders are situated in contexts that are too complex to easily quantify, this dissertation explored how stakeholders engaged with their domain expertise to make the judgement calls about data to support equity-oriented goals. This reflects a shift in perspectives about data from objective to socially-constructed, from from positivist to critical.

3. Unifying framework. To understand the process of interpreting data for equity in educational contexts, I identify three factors that contribute to the formations of beliefs and experiences when stakeholders interpret data for equity-oriented goals. This framework is shown in Figure 1. Expanding upon a framework for equitable data sensemaking [1], I described three factors that contribute to beliefs and experiences when interpreting data for equity: *Relevant prior knowledge*, *perceptions of power relationships*, and *cultural competence* [5, 29]. I frame interpreting data as connecting new knowledge to existing knowledge frameworks, so this framework considers prior knowledge that stakeholders deem relevant. At the foundation of my approach for supporting equity-oriented goals is a commitment to a critical perspective, which requires a consideration of power relationships in systems, ideologies, and institutions [26]. Stakeholders' relationships within social systems vary, so this framework considers how stakeholders interpreting data perceive power relationships relative to their own positionality [13]. Cultural competence is a model to guide actions taken at individual, organizational, and systemic levels to meet the needs of culturally and racially diverse groups in a culturally appropriate way [5, 29]. I consider cultural competence as a factor that affects how stakeholders interpret data. I posit that more developed cultural competence can support deeper engagement and interpretations of data. More cultural competence in interpreting data can consider the data as a signal of disparities or bias that arise from exclusive or harmful systemic policies and practice.

3.1 Contributions to computing education research and learning analytics. While data-driven techniques (e.g. statistical and machine learning) benefit from clearly defined dichotomies and independent features which purport to reflect all relevant information, stakeholders' identities are fluid, intertwined in intersectionality, and responsive to an everchanging context. That is to say that the quantitiative methods rely on static, isolated, and easily measureable features, but learners and their environments are dynamic, situated, and not easy to quantify [9]. This lack of alignment suggests that to support equity-oriented goals, we must design ways to enable direct stakeholders to interpret data by connecting it to their domain expertise. This has implications to learning analytics/learning at scale and computing education research communities.

Traditionally the learning analytics and neighboring research communities (e.g. Learning Analytics and Knowledge, Learning at Scale) has wielded quantitative methods in their use of data to understand and model learning experiences. But they have struggled to support equitable learning experiences [22] and "close the loop" by making data-driven insights actionable [3]. This dissertation makes steps towards using data for equitable action by contributing a framework and design explorations with empirical evaluations that inform the design of interactions with data that enable stakeholders to interpret data in support of equity-oriented goal.

Computing education research has used data to identify unequal disparities in participation and learning outcomes (e.g. [24]) and conducted investigations of equity where researchers were the interpreters of data (e.g. [20]). But it is still an open question how data may inform equity-oriented goals, which are more more socially situated. This dissertation contributed a framework and design explorations that considered a new approach to supporting equity-oriented goals in computing education that positions situated stakeholders as interpreters of data.

4. Codeitz: Affording and informing agency in online learning. After conducting a study to investigate the academic and social experiences of transfer students studying computer science at the UW Allen School [10], I identified how self-directed online learning tools were necessary for transfer students and other minoritized groups learning computing. But I also found these tools often failed to equitably serve them.

Typical self-directed online learning experiences tend to either 1) be too standardized to provide learners with personally relevant feedback and guidance (e.g. Khan Academy) or 2) too prescriptive such that learners lack agency and cede control of their learning experience to a data-driven system (e.g. Intelligent Tutoring Systems). They often fail to provide the information and context (via guidance, feedback, etc.) to inform a learner as well as afford agency so a learner has opportunities to guide their own learning experiences. Affording *agency*, or the sense a learner is in control of their actions and their effects, is critical to learning [15]. A resulting inequity is that online leaning environments are not inclusive to learners of varying self-efficacy, where high self-efficacy learners may desire the agency to guide their own learning, and low self-efficacy learners may desire guidance on what to learn next.

To explore how enabling alternative options that balanced agency and automation could support learners of varying levels of self-efficacy, I designed three versions of Codeitz, a self-directed online learning environment [32]. In collaboration with researchers from the UW Information School, Allen School, and College of Education, I designed three variations of Codeitz that each had different subsets of features that afforded or informed agency to varying levels. Figure 2 shows how different navigational features were included in the informed high-agency, uninformed high agency, and informed low-agency variations of Codeitz. In the two high-agency conditions, learners could decide what to learn next and choose for themselves. In the two informed conditions, learners had access to adaptive information from a predictive model we built that estimated knowledge levels and recommended next tasks based on their previous actions. All variations of Codeitz shared the same introductory Python curriculum that followed a theory of instruction for introductory programming skills that I had previously proposed and evaluated [31]. Codeitz enabled us to explore how informing and affording agency may equitably support learners of varying levels of self-efficacy.

In a study with 79 novice programmers, I found that that while varying agency and information in three separate conditions affected engagement, it did not yield differences in learning outcomes (Table 1). Furthermore, participants reported the predictive information Codeitz had (skill bars, recommendations) were less helpful than more typical feedback (concept overview, progress indicators, exercise correctness). Qualitative analysis suggested that learners wanted the flexibility to guide their own learning experiences when they wanted to and to cede the decision-making at other times. I interpreted these findings as design implications to suggest that expressing agency may deviate from the expectation of being told what to do, that perceptions of adaptive indicators evolve, and that affording agency requires considering the structure of the concepts to learn. By designing a tool that provides contextual information to inform decisions and opportunities for learners to exert agency without requiring it, I contributed guidelines for how to design for equitable self-directed learning by informing learners so they can balance agency and automated guidance.

Connecting the findings of this design exploration back to the framework I defined in Fig. 1, this study suggested that the design of Codeitz did not effectively consider students' *relevant prior knowledge* and *perceptions of power relationships* in regards to exercising agency in a learning experience. Most participants were familiar with formal higher education learning experiences, where instructional design often left little room for self-directed learning. So participants were likely unfamiliar with exercising agency, resulting in them disregarding data that could have informed their self-directed learning.

5. Differential Item Functioning (DIF) to detect assessment bias. Understanding inequity at scale is necessary for designing equitable online learning experiences, but also difficult. Statistical techniques like Differential Item Functioning (DIF) ([34]) can help identify whether items/questions in an assessment exhibit potential bias by disadvantaging certain groups (e.g. whether item disadvantages woman vs man of equivalent knowledge). While testing companies typically use DIF to identify items to remove, I explored how domain-experts such as curriculum designers could use DIF to better understand how to design instructional materials to better serve students from diverse groups. Using Code.org's online Computer Science Discoveries (CSD) curriculum, I analyzed 139,097 responses from 19,617 students to identify DIF (potential bias) by gender and race in assessment items (e.g. multiple choice questions). Of the 17 items, six disadvantaged students who reported as female when compared to students who reported as non-binary or male. I also identified that most (13) items disadvantaged AHNP (African/Black, Hispanic/Latinx, Native American/Alaskan Native, Pacific Islander) students compared to WA (white, Asian) students. Trace plots of items exhibiting DIF with medium or large effect size are shown in Figure 3. Through simulations, we found that students of equivalent knowledge levels but different genders and races would score statistically differently on this assessment, as shown in Figure 4. Through quantitative analyses, I was able to identify hidden patterns of bias within an assessment.

To translate data on bias to equitable action, I conducted a workshop and interviews with seven curriculum designers to identify how they interpreted data on test item bias. I found that they interpreted bias relative to an intersection of item features and student identity, the broader curriculum, and differing uses for assessments. I interpreted these findings in the broader context of using data on assessment bias to inform domain-experts' efforts to design more equitable learning experiences. This work contributed evidence to support a new use of DIF that connected data that identified nuanced biases in assessments with stakeholders who have domain expertise to take equitable action.

Our analysis of gender-based DIF data was the first to include reported non-binary students, but it also demonstrated how a typical positivist framing would lead to misinterpretation of the data. Analysis of the student response data and incomplete self-reported demographics data identified test questions that *positively* favored non-binary students. Without further context, the data-driven conclusion would be that non-binary students were positively biased when compared to reported male students. But when considered in the context of additional information on experiences of non-binary K-12 students (e.g. how coming out is a constant and challenging process [16]), we come to a more nuanced understanding of how the data may not reflect the experiences of non-binary students. This example reflects a need to shift from a postivist framing of considering data as unbiased and isolated to a more situated and contextual critical framing that instead problematizes data.

Connecting the findings of this design exploration back to the framework I defined in Fig. 1, this study suggested that curriculum designers were able to engage their *relevant prior knowledge* and *cultural competence* to contextualze data on potential test bias (DIF). The collaborative workshop provided designers with an opportunity to consider data in relation to prior knowledge they had with the design and use of the test and curriculum. And some curriculum designers with more cultural competence from their prior education and lived experiences were able to draw potential connections between potential test bias and systemic challenges that extend beyond curriculum design. But *perceptions of power relationships* scoped the conversation to what actions were feasible given their responsibilities and limited time and resources.

6. StudentAmp: Contextualized student feedback in large courses. As computing courses become larger and remote, online learning becomes more common, 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), which in turn can impact career aspirations and sense of belonging in a field. 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, I developed StudentAmp, an interactive system that collected contextualized student feedback at scale while ensuring student privacy. Using StudentAmp, students shared challenges they faced, demographic information (e.g. gender, ethnicity, disability, educational background), and perceptions of peers' challenges. Figure 5 shows how the contextualized feedback was shown to instructors and teaching assistants (TAs). Using StudentAmp, instructors and TAs collaboratively labeled challenges and then used filters to identify what kinds of challenges disproportionately affected which groups of students.

In collaboration with colleagues from the UW Allen School and Information School, I conducted an evaluation with five large (150 - 750 students), remote courses during the COVID-19 pandemic to understand how teaching teams used data on contextualized student feedback. We coded the 810 challenges students reported with StudentAmp into 17 themes that reflected challenges relating to the course that used StudentAmp (course structure, course content, remote learning), academic life (e.g. other classes, extra curriculars), non-academic roles (e.g. home family, job), broader context (e.g. geographic location, COVID-19, political unrest), and individual well-being (e.g. physical health, mental health, motivation). Through interviews with students from minoritized groups, we identified that students felt it was important to share challenges beyond the scope of the course, that demographic information was seen as an asset, and that seeing peers' challenges fostered a sense of community. From group interviews with instrutors and TAs, we discovered how cultural competence from prior training (e.g. coursework in public health, anti-racist seminars) enabled some to engage more readily with topics of identity and minoritization. We also discovered a tension between relying on prior knowledge of personal experiences and recognizing that lived experiences of students varied from their own. And finally, we uncovered how perceptions of power relationships left even professors feeling somewhat powerless to address systemic inequities that extended beyond their courses. These findings inform the design of contextualized student feedback that equitably considers the needs of students at scale while also ensuring the privacy and well-being of the minoritized groups.

Connecting the findings of this design exploration back to the framework I defined in Fig. 1, this study suggested *cultural competence* supported teaching teams' interpretations of demographic data to consider perspectives of minoritized groups. Instructors and TA had developed their cultural competence through additional training (e.g. seminars, coursework, research) demonstrated more capability to engage with the demographic data to consider perspectives of minoritized groups. This helped teaching teams shift away from an implicit assumption that all students had experiences similar to their own or that students aligned with dominant groups. And similar to the workshop with curriculum designers in the previous chapter, *prior knowledge* related to teaching and taking the course helped teaching teams contextualize challenges students reported, and *perceptions of power relationships* focused the conversation largely on factors that they could control.

7. Discussion & Future work. There are multiple ways to interpret the findings from these three studies as they relate to designing interactions to enable stakeholders to use data to support equity-oriented goals. A key theme across these interpretations is equity-oriented goals are situated in dynamic social contexts. Thus, I argue that we must embrace a plurality of approaches that are unified around a common commitment of problematizing the dominant social norms by questioning what we accept as normal in computing education and who we minoritize in the process. All this is in an effort to change dominant structures, systems, and discourses to imagine and work towards more equitable and just futures in computing education.

One interpretation of these findings is that the reductions and simplifications that come with data are harmful to equity-oriented goals. This is a common and valid critique amongst critical data scholars (e.g. [2, 6]), situated on a history of using quantitative methods to justify and perpetuate minoritization and harm [7, 6]. While data and associated quantitative methods are dangerous, they may not be inherently oppressive. In this dissertation, I argue data can support equity-oriented goals when interactions with data promote open interpretations and connections with other forms of knowledge. The intention of the design of Codeitz was to provide recommendations to provide learners with the opportunity to consider it relative to other information they had. In the workshop with DIF, I scaffolded the experience to have curriculum designers consider what actions they may take or what information they were missing when they interpreted data on potential test bias. Similarly for StudentAmp, I asked teaching teams to explore the data on student feedback and demographics and consider it in relation with relevant prior knowledge related to teaching. Combined, these studies explore the use of data as one of many tools by putting it in the hands of stakeholders who have the domain expertise to interpret it amongst other tools. Future work may explore in how to scaffold connections of data interactions with other knowledge to promote equitable action and dissuade unproductive deficit framings of minoritized groups [27].

Another interpretation of these findings is that we must design interactions with data that consider how stakeholders form prior beliefs that affect their interpretations. For these studies, I considered how stakeholders' cultural competence, relevant prior knowledge, and perceptions of power relationships affected their interpretations of data for equity. Across these studies, I identified how stakeholders engaged with their prior knowledge when interpreting data. We also identified how perceptions of power relationships focused much of the interpretations on factors that they could control (e.g. curriculum and test design for curriculum designers, course structure for teaching teams). This suggests promising future work related to multiple stakeholders interpreting the "same" data, perhaps in collaboration. And while these studies did not foster the development of cultural competence, they demonstrated how stakeholders with prior experiences relevant to cultural competence (e.g. coursework in public health, lived experience as part of minoritized groups) made interpretations richer by framing them within broader systemic contexts of minoritization.

Future work can elaborate on the framework to consider how to situate data for equitable action in new contexts. The framework I proposed (Fig. 1) was a result of design explorations that I conducted within the contexts of computing education. Computing education's challenges with inequities and minoritization of many groups also apply to other contexts, such as neighboring educational disciplines (e.g. science, math, engineering education) and perhaps other communities that engage with computing (e.g. technology companies, government agencies considering the use of technology). Future work can explore how this framework could apply to support equity-oriented goals in these contexts. Through further application of this framework to inform design of interactions with data, we can simultaneously explore how relevant prior knowledge, perceptions of power relationships, and cultural competence vary by context, time, and positionality of stakeholders. By doing so, we can understand how stakeholders' domain expertise enrich their interpretations of data for equity-oriented goals. It is through this situated, critical framing of data that we can use data to address inequities it more often perpetuates.

8. Conclusion. Considering data for equity-oriented goals can be uncomfortable or uncertain to stakeholders with their own prior knowledge, power relationships, and cultural competence. A lack of consideration of factors such as these can result in interpretations of data that are not situated in a social context. These acontextual interpretations will often fall short of supporting equity-oriented goals that are socially situated. Therefore, I conclude by reiterating my thesis statement, which reflects a broader call to consider the social, historical, political, and economic contexts surrounding data and the people who interact with it:

Interactions with data that consider prior knowledge, perceptions of power relationships, and cultural competency can enable computing education stakeholders to connect their interpretations of data with their domain expertise in service of equity-oriented goals.

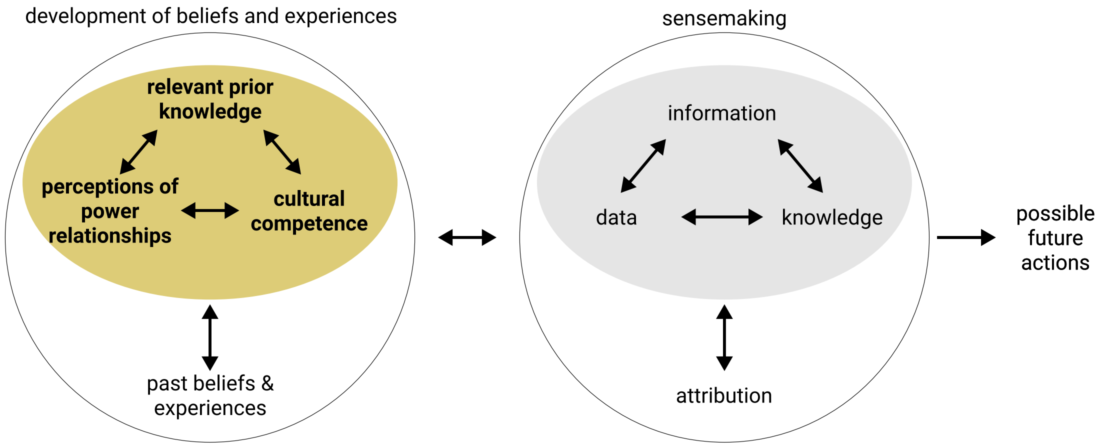

Figure 1: Unifying framework of my dissertation: The development of beliefs and experiences requires consideration of relevant prior knowledge, perceptions of power relationships, and cultural competence.

References.

[1] Melanie Bertrand and Julie A Marsh. Teachers' sensemaking of data and implications for equity. *American Educational Research Journal*, 52(5):861–893, 2015.

[2] Danah Boyd and Kate Crawford. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, 15(5), 2012.

[3] Doug Clow. The learning analytics cycle: Closing the loop effectively. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge*, LAK '12, pages 134–138. ACM, 2012.

[4] Kimberlé Williams Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. In Martha Albertson Fineman and Rixanne Mykitiuk, editors, *The Public Nature of Private Violence*, pages 93–118. Routledge, 1994.

[5] Terry L Cross, Barbara J Bazron, Karl W Dennis, and Mareasa R Isaacs. Towards a culturally competent system of care: A monograph on effective services for minority children who are severely emotionally disturbed. Technical report, Georgetown University, March 1989.

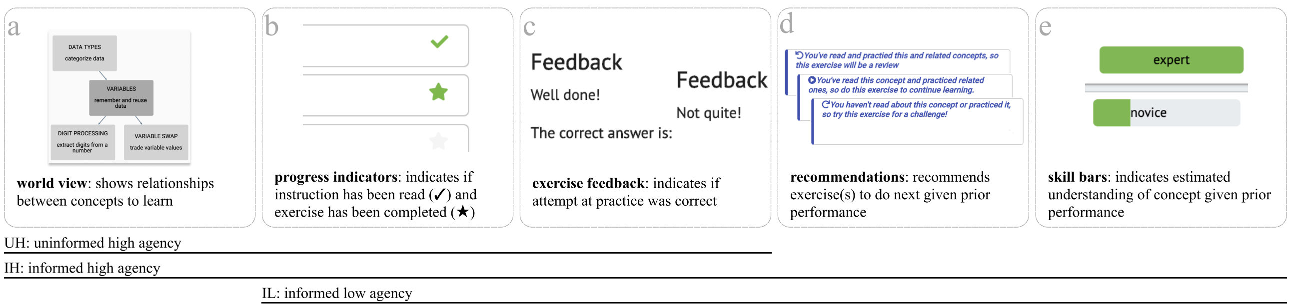

Figure 2: Features of Codeitz designed to provide learners with information for deciding what to learn next. Variations of Codeitz exposed learners to different subsets of the features (see lines at bottom of figure).

Table 1: Data by condition. Sample size (n) includes number of low (↓) and high (↑) performers on post-test. Histograms of post-test score (max: 39.5), number of Codeitz exercises attempted (max: 44), and number completed (max: 43) shown with median ($\tilde{x}$) and interquartile range (iqr) (approx. to histogram bin).

cond.	n	test score	# attempted	# completed
IH (Informed High Agency)	25 ↓: 7. ↑:9	$\tilde{x}$:23.5; iqr:12	$\tilde{x}$:43; iqr:1	$\tilde{x}$: 43; iqr: 3
IL: Informed Low Agency	31 ↓:12. ↑:12	$\tilde{x}$:21.8; iqr:17	$\tilde{x}$:29; iqr:34	$\tilde{x}$:29; iqr:34
UH: Uninformed High Agency	23 ↓: 7. ↑:5	$\tilde{x}$:23.0; iqr:14	$\tilde{x}$:43; iqr:0	$\tilde{x}$:43; iqr:2.5

[6] Alain Desrosières. *The Politics of Large Numbers: A History of Statistical Reasoning*. Harvard University Press, 2002.

[7] Hamid Ekbia, Michael Mattioli, Inna Kouper, G Arave, Ali Ghazinejad, Timothy Bowman, Venkata Ratandeep Suri, Andrew Tsou, Scott Weingart, and Cassidy R Sugimoto. Big data, bigger dilemmas: A critical review: Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66(8):1523–1545, 2015.

[8] Michel Foucault. *The Foucault Reader*. Pantheon Books, 1984.

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

[10] Harrison Kwik, Benjamin Xie, and Amy J Ko. Experiences of computer science transfer students. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*, ICER '18, pages 115–123. ACM Press, 2018.

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

[12] James D Marshall. Michel foucault: Educational research as problematisation. In *Why Foucault?: New Directions in Educational Research*, pages 15–28. Peter Lang, 2007.

[13] Peter McLaren. *Critical Pedagogy: A Look at the Major Concepts*. Routledge/Falmer Press, 2002.

[14] Paola Medel and Vahab Pournaghshband. Eliminating gender bias in computer science education materials. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education - SIGCSE '17*, pages 411–416, Seattle, Washington, USA, 2017. ACM Press.

[15] Sarah Mercer. The complexity of learner agency. *Apples - Journal of Applied Language Studies*, 2012.

[16] Kevin L Nadal. *The SAGE Encyclopedia of Psychology and Gender*. SAGE Publications, April 2017.

[17] Abelardo Pardo and George Siemens. Ethical and privacy principles for learning analytics. *British journal of educational technology: journal of the Council for Educational Technology*, 45(3):438–450, May 2014.

[18] Heather E Price. Large-Scale datasets and social justice: Measuring inequality in opportunities to learn. In Kamden K Strunk and Leslie Ann Locke, editors, *Research Methods for Social Justice and Equity in Education*, pages 203–215. Springer International Publishing, Cham, 2019.

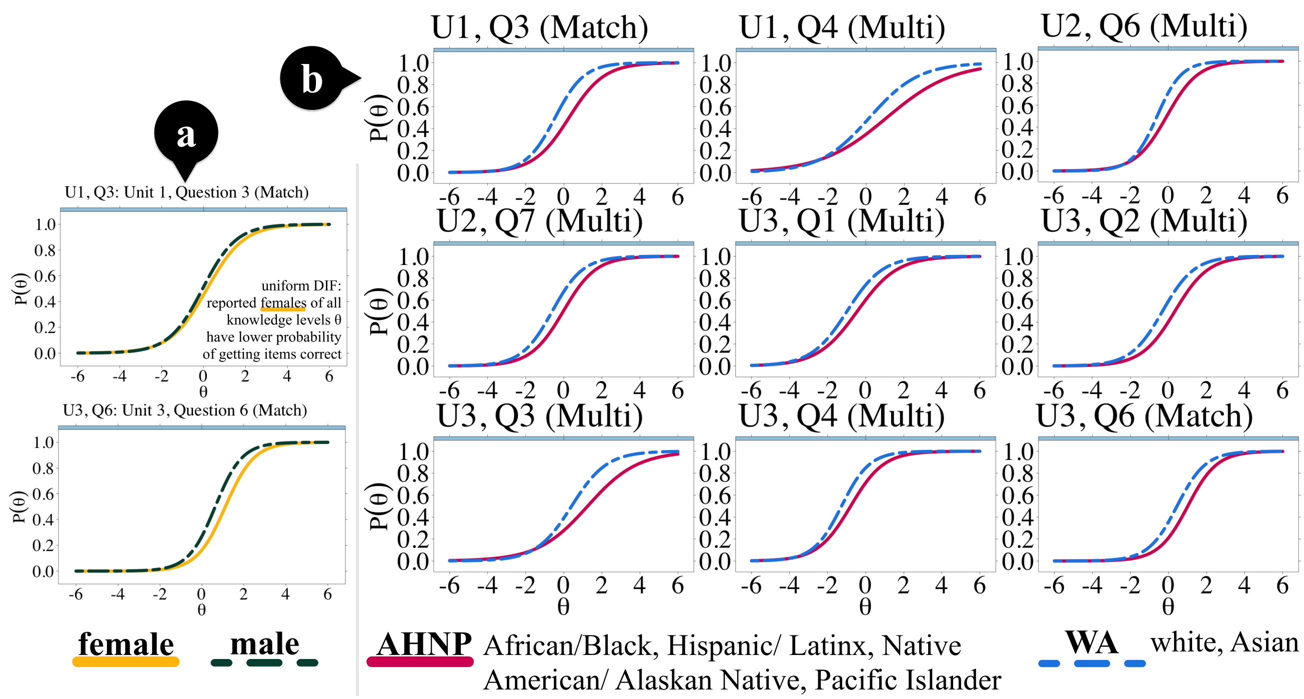

Figure 3: Traces for items that exhibited (uniform) a) gender-based DIF (medium or large effect size) and b) race-based DIF (large effect).

[19] Paul Prinsloo and Sharon Slade. Student vulnerability, agency and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1):159–182, April 2016.

[20] Yolanda A Rankin, Jakita O Thomas, and Sheena Erete. Real talk: Saturated sites of violence in CS education. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, SIGCSE '21, pages 802–808, New York, NY, USA, March 2021. Association for Computing Machinery.

[21] Justin Reich. *Failure to Disrupt: Why Technology Alone Can't Transform Education*. Harvard University Press, 2020.

[22] Justin Reich and Mizuko Ito. From good intentions to real outcomes: Equity by design in learning technologies. Technical report, Digital Media and Learning Research Hub, 2017.

[23] Kevin Robinson, Keyarash Jahanian, and Justin Reich. 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*, SIGCSE '18, pages 882–887, New York, NY, USA, February 2018. Association for Computing Machinery.

[24] Monique Ross, Zahra Hazari, Gerhard Sonnert, and Philip Sadler. The intersection of being black and being a woman: Examining the effect of social computing relationships on computer science career choice. *ACM Trans. Comput. Educ.*, 20(2):1–15, February 2020.

[25] Sharon Slade and Paul Prinsloo. Learning analytics: Ethical issues and dilemmas. *The American behavioral scientist*, 57(10):1510–1529, October 2013.

[26] Kamden K Strunk and Jasmine S Betties. Using critical theory in educational research. In Kamden K Strunk and Leslie Ann Locke, editors, *Research Methods for Social Justice and Equity in Education*, pages 71–79. Springer International Publishing, Cham, 2019.

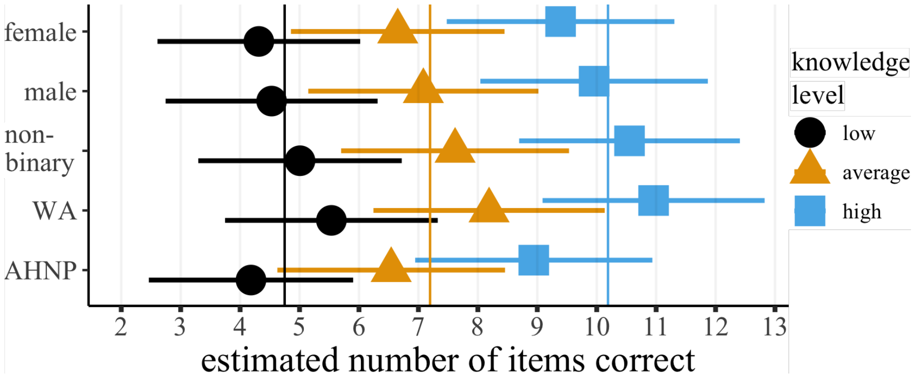

Figure 4: Expected number of items a student would get correct (out of 17) by gender and racial groups for three different knowledge levels. Knowledge levels were calculated with an Item Response Theory model assuming no DIF, where *average* is the median knowledge level in our sample ($\theta = -0.07$), *low* is a standard deviation (1σ) below ($\theta = -0.81$), and *high* is 1σ above ($\theta = 0.65$). Vertical bars indicate simulated mean number correct with no DIF. Shapes indicate mean number of items correct for each group from 1000 simulations, with horizontal error bars showing 1σ.

[27] Richard R Valencia. *The Evolution of Deficit Thinking: Educational Thought and Practice*. Routledge, November 2012.

[28] José Van Dijck. Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & society*, 12(2):197, 2014.

[29] Alicia Nicki Washington. 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*, SIGCSE '20, pages 213–219, New York, NY, USA, February 2020. Association for Computing Machinery.

[30] Benjamin Xie. How data can support equity in computing education. *XRDS: Crossroads, The ACM Magazine for Students*, 27(2):48–52, December 2020.

[31] Benjamin Xie, Dastyni Loksa, Greg L Nelson, Matthew J Davidson, Dongsheng Dong, Harrison Kwik, Alex Hui Tan, Leanne Hwa, Min Li, and Amy J Ko. A theory of instruction for introductory programming skills. *Computer Science Education*, pages 1–49, January 2019.

[32] Benjamin Xie, Greg L Nelson, Harshitha Akkaraju, William Kwok, and Amy J Ko. The effect of informing agency in Self-Directed online learning environments. In *Proceedings of the Seventh (2020) ACM Conference on Learning @ Scale*, L@S 2020, pages 77–89. ACM, 2020.

[33] Michael Zieky. Practical questions in the use of DIF statistics in test development. In Paul W Holland and Howard Wainer, editors, *Differential Item Functioning*, pages 337–347. Erlbaum, 1993.

[34] Bruno D Zumbo. Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language assessment quarterly*, 4(2):223–233, 2007.

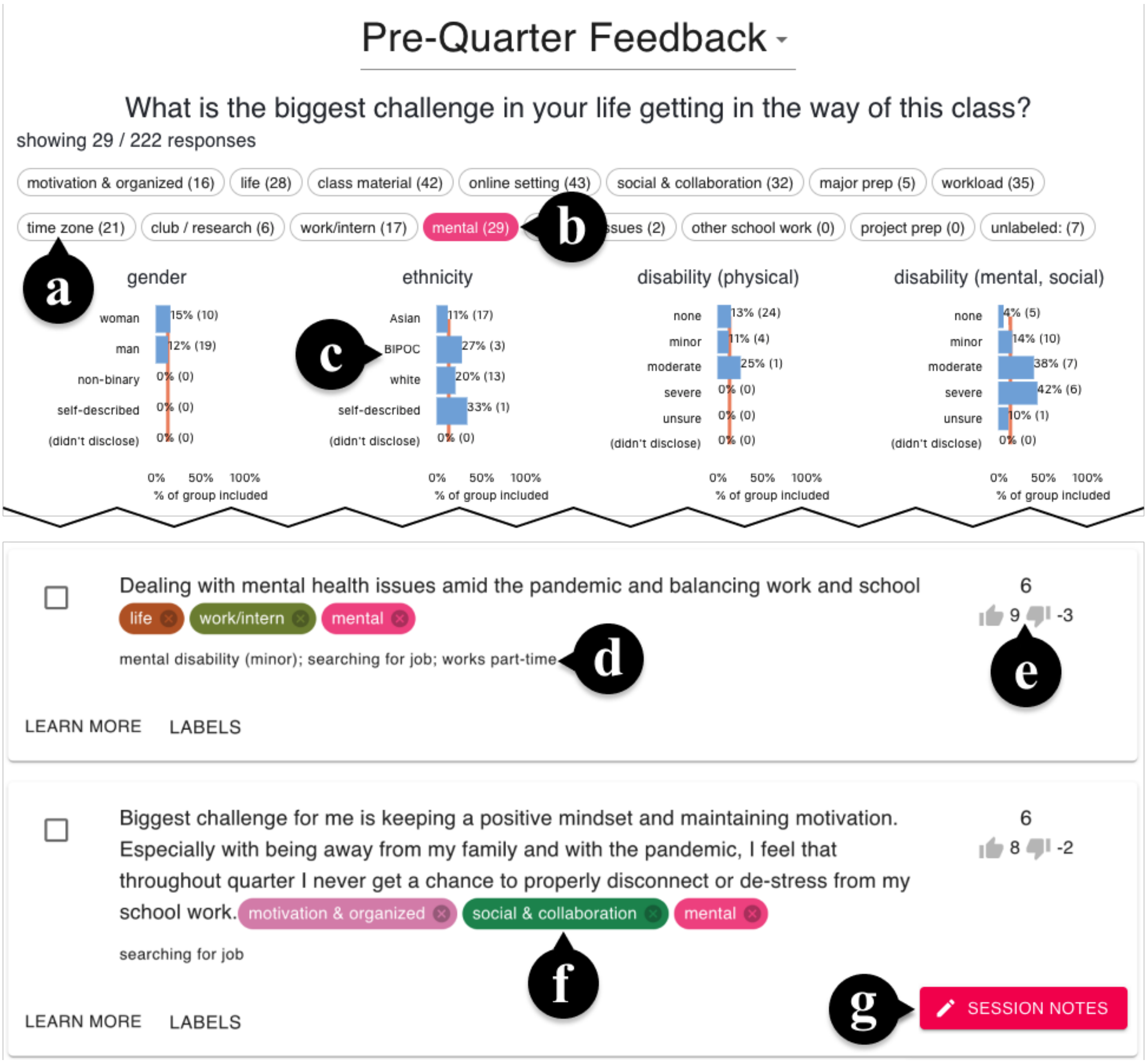

Figure 5: StudentAmp instructor view: Teaching teams could organize challenges by creating custom labels (a), which they could select to filter responses (b). The filters enabled teaching teams to use charts of demographic information (c) to see how challenges disproportionately affected certain groups (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.