Interpretations and Uses of Data for Equity in Computing Education

Benjamin Xie
bxie@uw.edu
University of Washington
The Information School, DUB Group
Seattle, Washington, USA

ABSTRACT

Computing education's booming enrollment exacerbates inclusion challenges ranging from tools that do not support diverse learners to instructors not being aware of unique challenges that students of minoritized groups face. While data often perpetuates inequities in many contexts, it could also serve to support equity-related goals if properly contextualized. To understand how data could support equitable learning, I explore how affording information and agency supports students' self-directed learning of Python programming, how contextualizing psychometric data on test bias with curriculum designers' domain expertise could support equitable curriculum improvements, and how contextualizing student feedback with demographic information and peer perspectives could help instructors become aware of challenges that students from minoritized groups face while preserving student privacy and well-being. By studying how students, curriculum designers, and teachers interpreted and used data relating to experiences learning computing, I contribute techniques that contextualize equity-oriented interpretations and uses of data with stakeholders' domain expertise.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**; *Student assessment*; • **Human-centered computing** → *Human computer interaction (HCI)*.

KEYWORDS

computing education; equity; student feedback; assessment

ACM Reference Format:
Benjamin Xie. 2021. Interpretations and Uses of Data for Equity in Computing Education. In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3446871.3469780

1 INTRODUCTION: INEQUITIES IN CSED

Computing education simultaneously faces booming enrollment as well as persistent inclusion issues. Post-secondary degrees in computing (e.g. bachelor degrees in computer and information sciences) are becoming more required for participation in the computing workforce, so undergraduate computing degrees are important to those interested in working in the computing space. In part due to the growing demand for computing skills, enrollment numbers for computing degrees has surged recently, straining resources and educators [3]. But computing classes still face similar diversity issues as computing and STEM workforces [3].

To address inclusion issues in computing education, we must enable more equitable learning experiences. I frame equity as referring to access to and successful participation and achievement in learning computing. Equity has a social justice goal where corrective measures must adjust for aggregate harm from social inequalities [4]. Also, equity cannot exist without first understanding inequalities relative to some baseline (e.g. proportion of women in computing workforce vs proportion of women in population) [4]. Approaches to make education more equitable include using data to inform instructors of their students' needs (e.g. [5]).

With a growing reliance on data-driven technology to support learning experiences comes greater concerns of equity. The enrollment boom in computing education pressures educators to use technology to standardize learning experiences and assessments. Standardized learning experiences tend to be disconnected from the lives of students from minoritized groups, potentially resulting in dangerous unintended consequences for the issue of equity [1]. Personalization through the use of data-driven tools such as Intelligent Tutoring Systems (ITS) attempt to automatically personalize learning experiences (e.g. using prior response history to prescribe a next practice exercise), but these tools are typically trained on data that does not consider demographic differences.

I designed tools and techniques that contextualized data to support equitable action. Critical theorists have framed data as a tool for oppression and traditional statistical techniques may explain away unique experiences of student from minoritized groups as noise or outliers. I investigated into an alternate perspective, where we can enable stakeholders can take timely, informed, equitable action by interpreting data within the context of their domain expertise.

2 CODEITZ: AGENCY IN ONLINE LEARNING

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.

ICER 2021, August 16–19, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8326-4/21/08.
https://doi.org/10.1145/3446871.3469780

Affording *agency*, or the sense a learner is in control of their actions and their effects, is critical to learning [2].

To explore how enabling alternative options that balance agency and automation can support learners of varying levels of self-efficacy, I designed three versions of Codeitz, a self-directed online learning environment. When a student learning Python has to decide what to learn next, they can choose for themselves or follow recommendations from a system trained on their prior actions [6]. These versions compared how varying levels of agency and amount of predictive information affected learning outcomes [6]. I found that that while varying agency and information in three separate conditions affected engagement, it did not yield differences in learning outcomes. 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 learning by informing learners so they can balance agency and automation.

3 INTERPRETATIONS OF TEST BIAS

Understanding inequity at scale is necessary for designing equitable online learning experiences, but also difficult. Statistical techniques like Differential Item Functioning (DIF) ([7]) 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 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. I then conducted a workshop and interviews with seven curriculum designers and found that they interpreted item 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.

4 STUDENT AMP: EQUITABLE FEEDBACK

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 Student Amp, an interactive system that used student feedback and self-reported demographic information (e.g. gender, ethnicity, disability, educational background) to show challenges and how they may disproportionately affect certain groups. To help instructors make sense of feedback, Student Amp ranks challenges by peer perceived disruptiveness of challenges. I conducted formative evaluations within five large college computing courses (163 - 628 students) to understand how using Student Amp affected instructors' perceptions of their students' experiences. We found that by enabling inclusive, efficient, and scalable student feedback, Student Amp was able to inform instructors and students to develop a collective awareness of inequities in computing courses, fostering empathy and understanding amongst instructors and students and informing broader conversations on addressing course-level and systemic challenges. We interpreted our results as they related to considering risks and assets when incorporating demographic data as context to feedback systems that advance equity-related goals.

5 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1735123, 12566082, 2031265, 1703304, 1539179, a Google Cloud Academic Research Grant, and gifts from Microsoft, Adobe, and Google. All studies were preapproved by the UW Institutional Review Board (IRB).

REFERENCES

[1] Paulo Blikstein and Sepi Hejazi Moghadam. 2019. Computing Education Literature Review and Voices from the Field. In *The Cambridge Handbook of Computing Education Research*. Cambridge University Press, 56–78.

[2] Sarah Mercer. 2012. The Complexity of Learner Agency. *Apples - Journal of Applied Language Studies* (2012).

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

[4] 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.

[5] Niral Shah, Daniel Reinholz, Lynette Guzman, Kenneth Bradfield, and Gregory Beaudine. 2016. Equitable Participation in a Mathematics Classroom from a Quantitative Perspective. In *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. 7.

[6] Benjamin Xie, Greg L Nelson, Harshitha Akkaraju, William Kwok, and Amy J Ko. 2020. 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)*. ACM.

[7] Bruno D Zumbo. 2007. Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Lang. Assess. Q.* 4, 2 (2007).