Hi, my name is Benjamin Xie
I’m a PhD candidate at the University of Washington Seattle,
a university which acknowledges the Coast Salish peoples of this land, the land which touches the shared waters of all tribes and bands within the Duwamish (doo-amish), Puyallup (pee-all-up), Suquamish, Tulalip (too-lay-lip) and Muckleshoot nations.
And I’m EXCITED to defend my dissertation on “stakeholders’ interpretations of data for equitable computing education”
(TOO LONG: ~2 min)
Let’s begin with a little story about the relationship between data and reality. In the 1930s, General Drafting Co was creating a road map of New York state. To prevent anyone from copying their maps, they created a fictitious place called “Agloe.” The idea was that if anyone else produced a map with “Agloe” on it, they could sue for copyright infringement.
So data, the map in this case, is a representation of reality, created by General Drafting Co for the purposes of preventing copying. Fast forward 20 years and sure enough, competing company Rand McNally produced a map that included “Agloe.” When General Drafting Co tried to sue, Rand McNally lawyers defended themselves by saying Agloe actually DID exist.
**CLICK**
Because someone had seen Agloe on a map, realized nothing was there, and built the Agloe General Store.
**CLICK**
So not only is data a representation of reality, this map (the data) created a reality of its own with the creation of the Agloe General Store! And you may think that “ah yes, we as a society were so silly and nieve back in the day; something like this could never survive in our present information age.” But while the Agloe General Store has been closed for decades...
**CLICK** Agloe appeared on road maps as recently as 1990s and on Google Maps in 2014.
This duality between data being one of many representations of reality and also creating realities of its own is a critical framing that I will come back to throughout my dissertation.
So computing education is the context my dissertation actually explores:

Over the past few years, the interest in learn computing has grown at an incredible rate!

At my home institution, interest in majoring in computer science grew five fold over the past 10 years.

**CLICK** But despite this, equity gaps still exist. For example, the percentage of computer & information science majors who reported as female peaked in 1985 at less than 40% and currently is around 25%.

So more people want to learn computing, but there are issues in how we teach computing that make the learning exclusive to many groups.
And to understand why groups are excluded, we need only to look at what learning computing looks like in the United States:

By high school, many students have heard about computing. They’ve heard it gets you high paying jobs. But most high schools schools don’t offer any computing courses, such as AP computer science principles.

**CLICK** About half of students who earn bachelor’s degrees in computer and information science in the US spend some time taking courses at a 2 year university.

**CLICK** And those able to transfer to a four year university will rely on online support that is often inadequate, be measured by biased instruments, and generally experience feelings of exclusion.

So put simply, learning computing is not an equitable experience. But we can do better!

**CLICK** For my dissertation, I will focus on three equity issues: a lack of adequate online support for students, bias in the tests we use to measure student learning, and the lack of awareness of students needs that can result in more exclusionary learning experiences. There are many others, but these three exist in across many formal learning experiences.
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.

- 3 factors
- stakeholders (students, instructors, curriculum designers) interpreting
- equity-oriented goals: inclusive online learning, addressing bias in tests, ensuring instructors are aware of students’ needs

Here is the thesis statement I will spend the next 40 minutes proving to you. We’ll come back to this statement multiple times to unpack it, but I wanted to get you all thinking about this right away. **CLICK**
So to understand my dissertation requires me to clarify two related concepts, with the first being equity. Equity is about not just access to computing education, but also successful participation and achievement within it. **CLICK** And equity serves a social justice goal of being a corrective measure for aggregate harm. That is to say that equity is not about treating all students equally, but rather providing unique support to students so they have equal opportunities to succeed. **CLICK** And finally, understanding disparities or inequalities relative to a baseline can help us identify potential inequities. For example, less than 30% of CS majors are women but women make up about half of the population, providing strong evidence of systemic inequities in the learning experience. At a high level, we can say that equity-oriented goals are very situated and contextual!
And learning experiences are not equitable to students in _minoritized groups_. These groups are typically underprivileged, stigmatized, and disadvantaged at a _systemic level_.

Within computing education and for my research, groups can be minoritized by gender, ethnicity, language, disability, as well as prior educational privilege. In contrast, dominant groups are those who are typically privileged, unstigmatized and often thought of as “the norm.” They include white and Asian men who went directly from high school into a 4 year university and whose parents were college educated.
How has data been used in computing edu?

computing edu: data to identify disparities/inequalities in access, experiences, and achievement

eample: Dr. Monique Ross and colleagues analyzed survey responses to understand experiences of computing students who were Black women compared to non-Black women and Black men.

Identifying disparities is important; but unclear what you do with that information.

*CLICK* a neighboring field of learning analytics has explored the use data for data-driven adaptations (personalized learning... FUNGUS). This iterative cycle begins with data collection from learners to develop models and metrics, which inform interventions, which are supposed to benefit learners.

*CLICK* A critique of this field is that it is too fixated on the “data to metrics” part, and there is a lack of “closing the loop” to use data to inform interventions that benefit learners. A common explanation for this shortcoming is that interventions are contextualized, and data often lacks that rich context.
So at a high level, we can say that we can use data to identify nuanced patterns. But equity-oriented goals are very situated. Gap: data lacks context but equity requires context. How do we provide context to interpretations of data? (because how we interpret and make sense of data is also a very situated activity)
An existing framework to interpret data for equity

To understand how to interpret data for equity...
Bertrand & Marsh developed a theoretical framework for explaining how teachers interpret data for equity. Just as students come into classes with prior beliefs and experiences, people who interpret data do so based on beliefs and prior experiences. They identified how beliefs and past experiences of people interpreting data affect how people make sense of data to determine possible future actions to support equity. But they do not describe what factors contribute to the formation of beliefs and past experiences.

*CLICK*
Equity is a situated goal, so the beliefs and past experiences that somebody situates their interpretations of data in is absolutely critical! If we don’t consider this, we risk people disregarding the data or misinterpreting it in ways that can be harmful. Not because they are nefarious or a RAPSCALLION…
For my dissertation, I identified three factors that affect the formation of beliefs and experiences:
- relevant prior knowledge
- perceptions of power relationships, and
- cultural competence
(unpack)
Factors impacting interpretations of data for equity

prior: people interpret data relative to prior knowledge they deem relevant; connect to existing knowledge

**CLICK** cultural: 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. Four skills:
- Attitude: valuing how all factors of diversity are critical for an inclusive environment
- Awareness: recognition of own beliefs and positionality and how they interacts with others’
- skills: understanding historical impact of certain actions, words, beliefs and adapting to better meet needs of minoritized groups
- knowledge: institutionalized cultural knowledge across all organization levels

Development across these four skills range from cultural destructiveness to cultural proficiency

**CLICK** power relationships: Foucault and critical data studies. Relational, in systems, ideologies, institutions in a given context

Combined, I argue in my dissertation that these factors are critical to how people make sense about data for equitable computing education
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.

My thesis statement is this:
For my dissertation, I focus on three direct stakeholders: students learning computing, teachers providing instruction, and content designers who create curriculum and tests that students and teachers rely on.

My dissertation has three main projects:

**CLICK**
In my first project, I explored how to provide more equitable support for online learning by affording and informing agency.

**CLICK**
In my second project, I explored how content designers could use their domain expertise to contextualize test bias.

**CLICK**
And for my third project, I investigated how to contextualize student feedback to identify inequities in large remote courses.
the first project in my dissertation explores how to design self-directed online learning that supports that enables agency
context of self-directed online learning
often alone (without peers or instructors to support)
navigate their own experiences
Make decisions, take actions towards learning-related goals
Experience defined by how we design tools
how do learners navigate online experience?

learner decides
learner makes decisions informed by system
system decides

+: agency to explore
-: lack of info to guide
e.g. Coursera, Khan Academy

+: adaptive content
-: lack of agency
e.g. intelligent tutoring systems, adaptive tests

paradigms

learner decides
- massive open online courses, popular tools such as Khan Academy
- everything is there; they decide how to use it
- lack of guidance

on the other hand, system decides
- e.g. adaptive learning tools or intelligent tutoring systems
- adapts based on your prior actions
- not in charge of own learning experience

More informed agency
How does varying information & agency affect self-directed online learning?

A learner can take informed actions that align with their goals (Wardrip-Fruin et al. 2009)

proximal and action-related info key to making decisions
(Bettman, Luce, & Payne 1998; Lichtenstein & Slovic 2006)

interaction of information and agency
Critical to agency is decision-making
Design of interface > information > agency > learning outcomes
I wanted to explore interaction between agency and information to guide decision-making
Designed 3 variations of online learning tool (demo in next slides)
agency: low and high
information (adaptive info by system via BKT): uninformed, informed

Let me demonstrate the experience of using a variation of Codeitz
similar to the previous UH version
but info based on system predictions (Bayesian Knowledge Tracing)
adults, most of whom enrolled in post-secondary degree
participant feedback on importance, role of feature
- world view: high-agency only
- progress: useful across all conditions
- exercise: generally helpful, wanted more hints and feedback to fix mistakes

info based on system predictions (only for informed conditions)
- rec: least helpful of the features (paper)
- skill bars: move on or not
test scores: no difference across conditions

Potential explanations:

• most learners finished all exercises
• learners did not exercise agency
• assessment did not measure well

See paper for more explanations and qualitative data about this
- no agency: used to following instructions at undergrad studies
prior programming experience, greater self-efficacy predictive of higher test scores
As expected
results summary

- conditions had no effect on learning
- self-efficacy, prior knowledge had effects
- high-agency (IH, UH) did more practice
- skill bars, recommendations perceived as less important

- To summarize our results
- conditions w/ variations in information and agency afforded did not have detectable effect
- High agency did have more practice (may be indicator of motivation difference? More in paper)
- Skill bars, recommendations, info based on system predictions least helpful
design implications: agency is nuanced

perceptions
of adaptive indicators evolve

programming is unique domain

expectations:
agency may be unusual

“the order [of concepts] did not seem intuitive”

- recommendations:
  - Trust in recommendations is earned!
  - unintended interpretations, lack of trust or diminishing trust in adaptive feedback (cold start)
- domain: need to consider structure of domain. strict dependencies
  - ex: learning if/else before variables and relational operators may result in unproductive struggle
  - Think about what we want to design agency for
- expectations: may not be want to, comfortable, realize guiding own learning experience
<table>
<thead>
<tr>
<th>study</th>
<th>key finding</th>
<th>prior knowledge</th>
<th>perceptions of power relationships</th>
<th>cultural competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codeitz</td>
<td>learning not affected by agency</td>
<td>agency unfamiliar, deviated from expectations</td>
<td>(did not consider)</td>
<td></td>
</tr>
</tbody>
</table>
Domain experts’ interpretations of test bias

**Codeitz**
inform agency in online learning
agency must be an informed option

**DIF**
contextualizing test bias w/ domain expertise
data on bias requires judgement of domain experts

Domain Experts’ Interpretations of Assessment Bias in a Scaled, Online Computer Science Curriculum
Benjamin Xie, Matt J. Davidson, Baker Franke, Emily McLeod, Min Li, Amy J. Ko
LRS 2021

must design information in interfaces to enable agency
We use test scores for a lot of things. University use scores for test such as the AP CS exams to determine if a student should be accepted into a university or major. Teachers use tests for summative purposes such as grading. Students use tests to self-assess what they know, and the results can affect their self-efficacy and sense-of belongingness.

How do we know how good our tests are?

How different people interpret and use test scores is important, but tests are imperfect measures of knowledge.
To understand how good a test is, we have to make a few assumptions. Following Item Response Theory (IRT), we can assume that student knowledge and question difficulty are on the same continuous dimension. Say we have questions A, B, C, where A is the least difficult question and C is the most difficult. Say two students, Michelle and Jorge, answer these questions. Based on their responses to these questions to estimate their knowledge levels. Michelle's is between B and C because she got B correct but C incorrect. By a similar logic, Jorge's knowledge is between A and B.
Now say we wanted to look at a group of students who reported as male and a group who reported as female. All students got questions A and C correct. We would expect them to get question B correct as well. And say all the students who reported as male do get B correct.

**CLICK**

But say we observe that most students who report as female get question B wrong.

**CLICK**

This is a toy example that demonstrates Differential Item Functioning (DIF)

**CLICK**

where students of similar knowledge levels but different groups (genders in this case) perform differently on an item, question B in this case.

DIF is a technique to identify potential bias in test questions.
What if DIF could signal opportunities for better pedagogy?

But rather than use DIF as a filter, what if we could use it to improve equitably we teach?
Data on DIF can help identify or substantiate nuanced patterns of disparities or bias. But we need the domain expertise of stakeholders such as curriculum designers to interpret and use these findings to address inequities.
How do domain-experts use data on test bias by gender and race for equity?

So this work explores how domain-experts (curriculum designers) might be able to interpret and use gender and race-based DIF for equity related goals.
For this study, I partnered with Code.org, a nonprofit dedicated to inclusive. Computing education. We analyzed responses from ~20,000 students for middle school CS Discoveries (CSD) curriculum.
The 17 questions I analyzed were either matching questions or multiple choice questions. Matching questions required students to place options in their correct locations. Multiple choice questions required students to choose one or two options.
Our quantitative analysis focused on checking for potential test bias by gender and race.

**CLICK**

We say a question is biased if on average, a student from a disadvantaged group is at least 5% less likely to get that question correct compared to a student of similar knowledge from the other group. This is equivalent to checking for statistical significance with a medium or large effect size.

**CLICK**

We found two questions disadvantaged students who reported as female compared to students who reported as male.

**CLICK**

The figure on the right shows how for a test questioned that exhibited DIF…female… lower probability of getting question correct…compared to male

**CLICK**

Most test questions disadvantaged AHNP students (African/Black, Hispanic/Latinx, Native/Alaskan Native, and Pacific Islander) compared to WA students (white, Asian)
Put together, we can say that students of equivalent knowledge but different genders or races would score differently on the CSD assessments.

So as a whole, this test disadvantages AHNP and reported female students the most, and advantages WA students the most.
DIF does not tell us the cause of this bias or what to do about it. So to understand that, we conducted a remote workshop where 7 Code.org curriculum designers interpreted DIF data. All this was in an effort to understand a new use for DIF: improving equity in learning by informing domain experts of potential issues. Here are a few high level takeaways, but I point you to the paper to read more about our findings.
Curriculum designers considered how test design may have introduced bias, with some identifying how matching type questions disadvantaged students who reported as female.
Alignment between assessment and curriculum

“Comments are not very well emphasized in CS Discoveries… this may be the very first [time] that students are seeing this idea of putting a comment to a block of code.”

Curriculum designers also considered how the curriculum may or may not have prepared students for the test questions. So in one case, curriculum designers acknowledged that commenting code was a skill worth learning, but may not have been well taught prior to this test question. (explain example)
Considering how specific aspects test and curriculum design may contribute to bias is a potential first step to making changes that support more equitable learning experiences.
Iterating towards more equitable learning experiences requires measuring factors we cannot easily intuit, and using domain expertise to contextualize these findings with understanding we cannot easily measure.
Data helps us identify existence and extent of biases. Domain expertise helps us identify causes, take equitable action.

But the main take away is this:
study | key finding | prior knowledge | perceptions of power relationships | cultural competence
---|---|---|---|---
Codeitz | learning not affected by agency | agency unfamiliar, deviated from expectations | (did not consider) |
My work analyzing DIF with Codeorg curriculum designers demonstrated how data required judgement to act upon. And stakeholders who have the domain expertise to interpret this data to support equity-oriented goals.

For my third project I’ll share today… equitable student feedback… contextualizing… feedback on what students provide… information about who students are breaking news: this work was accepted to CSCW 2022 as of yesterday!
Here's the motivation for this work:

In higher education, teaching teams for large computing courses typically consist of a single instructor and a team of up to a few dozen student teaching assistants (TAs). And they have the responsibility of teaching anywhere from 150 - 650 students.

So if this small but dedicated teaching team wants to equitably support students, they would need to know what challenge students of minoritized groups were facing. But this can be quite difficult…especially remote…!

**CLICK**

When 1 in 4 students are women or non-binary students who risk potential stigmatization if they speak up for themselves, how can the teaching team know that keeping up with coursework is difficult because of some women students’ familial responsibilities at home?

**CLICK**

When 1 in 10 students are African/Black, Hispanic/Latinx, Native-American/Indigenous and Pacific Islander, how is a teaching team supposed to know that some of them have trouble getting the help they need to understand how code works because of the remote structure of the course.

**CLICK**

When 1 in 4 students transferred from another university are less familiar with norms of this university, how does a teaching know that some transfer students took the prerequisite coursework over two years ago and need to relearn concepts, when a vast majority of students took that coursework last term?

Put simply, students of minoritized groups face challenges that students of dominant groups (the majority of students) don’t face. Leaving these needs unknown and unmet is a major contributing factor to the inequities in classes. It’s not necessarily that teaching teams have ill-will or are ragamuffins; they often have to make assumptions about what students need based on their prior experiences or what they know about students, information that is biased towards those who are privileged.
enough to speak up and get noticed.

So this work explores how to inform teaching teams of needs of minoritized groups by amplifying their voice while also ensuring their privacy and wellbeing.
contextualized: challenges students face don’t occur in a vacuum. It’s a unique human being facing this challenge! A conversation between student and instructors can provide this context, but that’s not scalable and introduces potential social desirability biases
scalable: online form UNLABORIOUSLY
protect students: students shouldn’t have to risk potential stigmatization to advocate for themselves!

To create this equitable student feedback, I designed StudentAmp that provides contextualized, scalable, privacy-protecting student feedback.
What's the biggest challenge in your life getting in the way of the class?

“CLICK” Demographic information (intersectionality, multidimensional perspective taking)

“CLICK” Look at pairs of challenges that peers reported and determine which challenge was more disruptive.

Using a Copeland method for rank-choice voting, we aggregated these pairwise comparisons into a “disrupt score” that was shown to instructors. (more on that in a moment)
here's what the instructor would see after students share feedback using Student Amp:

see results in feedback session
challenges
demographics
disrupt score
label challenges
see how labels disproportionately affect certain groups
workload
BIPOC, work full time, transfer, moderate mental/social disability
1. What challenges do students share?
2. How do students perceive the values and risks of sharing?
3. How do teaching teams use contextualized feedback for equity-oriented interpretations?
at large research university in urban, tech-rich environment
data: St. Amp. responses + interviews

- data from Student Amp
  - challenges reported
  - demographics
  - feedback on peers’ challenges

- interviews
  - w/ students
  - w/ teaching teams
RQ1: themes from 810 challenges

inductive thematic analysis with subsequent round of qualitative coding using themes from initial analysis. 17 themes to represent 810 challenges.

course-related feedback: things closely related to course; often asked about in other feedback

**CLICK** remote learning: lack of structure with online classes (no walking between classes)

**CLICK** external to course. academic life (other classes, extracurriculars), non-academic life (familial, job), environment or broader context

**CLICK** home & family: BIPOC first generation women had to take care for her sister while father away

**CLICK** wellbeing: health, being isolated, struggles with self-regulation

**CLICK** mental health: depression and anxiety relating to everyday pressure

**CLICK** other: not reporting a challenge, or too vague

combined: students shared challenges beyond immediate scope of course
RQ2: How students perceived sharing

Challenges beyond the scope of the class were worth sharing

"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... can give you general knowledge of how it could be impacting the school bubble."
- S-D-36, BIPOC first-generation woman, works part-time, minor physical disability

"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."
- S-D-57

"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.'"
- S-A-148

Seeing others' challenges fostered community

Demographic information was seen as an asset

demographics, perceptions of peers' challenges
two rounds of interviews with 5 students from minoritized groups

"CLICK" challenges beyond scope of course worth sharing. BIPOC first generation woman: school was one of many bubbles.

"CLICK" demographic information as asset: good instructor understands structural injustices

"CLICK" seeing others' challenges: recall... determine disruptiveness... not the only facing this right now... belongingness during isolation
Moving on to how teaching teams interpreted contextualized feedback.

teaching teams were able to consider the challenges contextualized w/ demographics. Here’s an example of a discussion that a teaching team had:

challenge: woman, transfer student, severe mental disability. “I’m unsure of my ability to train my brain to think this way”

**CLICK** TA considered challenge relative to experiences taking and teaching

**CLICK** professor used demographics to consider multiple dimensions of student, drawing upon cultural competency from research in inclusive computing education

example of how prior knowledge and cultural competence helped interpret data
instructors also felt limited by systemic power structures
while much of the challenges focused on changes to course structure, likely because that's what teaching teams could control
equitable feedback involves considering positionality of students while ensuring wellbeing

not just assuming all students are the same, from dominant groups
<table>
<thead>
<tr>
<th>study</th>
<th>key finding</th>
<th>prior knowledge</th>
<th>perceptions of power relationships</th>
<th>cultural competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codeitz</td>
<td>learning not affected by agency</td>
<td>agency unfamiliar, deviated from expectations</td>
<td>(did not consider)</td>
<td></td>
</tr>
<tr>
<td>DIF/ test bias</td>
<td>designers identified potential changes to curriculum, test design</td>
<td>drew upon knowledge of curriculum, test design</td>
<td>focused on what they could control</td>
<td>varied, but enabled broader consideration of bias</td>
</tr>
<tr>
<td>StudentAmp</td>
<td>teaching teams considered how challenges affected subgroups differently</td>
<td>drew upon prior experience taking and teaching course, at institution</td>
<td>focused on changes within course</td>
<td>prior training enabled deeper consideration of student identity</td>
</tr>
</tbody>
</table>
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.

Here is the thesis statement I will spend the next 40 minutes proving to you. We'll come back to this statement multiple times to unpack it, but I wanted to get you all thinking about this right away.
Contextualized feedback enables consideration of minoritized perspectives.

Agency must be an informed option.

Codeitz: Informed agency in online learning.

Student Amp: Contextualizing student feedback.

DIF: Contextualizing test bias with domain expertise.

Data on bias requires judgement of domain experts.

Designers, teachers, students.
Just before I conclude, I want to very briefly acknowledge some of the beautiful humans who help me learn, grow, and thrive throughout my PhD:

- To my girlfriend Nicole: I’m difficult. Our PhDs are challenging. And it’s frustrating because I want to follow my instinct of working and making myself more work. But you bring me back to a reality where my life is more than what I do in front of a keyboard. So thanks for leaning in every time life became overwhelming. And thanks for convincing me to adopt our rescue dog Curie. She’s the highlight of WFH life.

- To my mom & dad: I’m not the kid who moved out for college a decade ago. And I don’t think my research is exactly what any of us expected me to do w/ my CS degrees. But I love you all for your unwavering support as we constantly try to figure out who I who I want to be.

- To the rest of my family: Thanks for showing me since childhood that me humans just being humans is pretty damn good.

- To my advisor Amy: I decided to do a PhD because I thought you would support me as a human being. I just didn’t realize how a single mortal human could provide such insight, mentorship, and support for me, our labmates, and so many other communities. And the while demonstrating how to be true to yourself.

- To my labmates: I get paid to collaborate and learn with you all, and that’s special. This PhD is characterized by the struggle, triumphs, and bewilderment I get to share through our problematization of every institution we’re aware of, birth of new ideas in whiteboard sessions, and Slack backchanneling.

- To the iSchool and DUB communities: Thanks for not only accepting me and my amoeba-shaped research ideas, but also putting perhaps blind faith in me to organize and run events. Deep down, I just like connecting people so we can have a shared experience, so thank you for the iSchool and DUB communities for letting me do that.

- To the mentors I’ve had along the way: Thanks for not only sharing knowledge with me, but also sharing such infectious excitement, all the while keeping things candid and real.

- To my friends: I’m sorry I’ve been dodging messages and bailing on hangouts recently. Thanks for your never-ending nudging to have me join your adventures.

- And finally: To everyone who works in the background. People like Dora who magically replenish my bank account after conferences, to folks in iSchool IT for getting me another laptop after my first one mysteriously disappeared, to folks in the IRB office who ensure my research does no harm, to Dr. Salazar in the UW counseling office for helping me whenever times felt too turbulent. UW functions because of your humility and dedication.
Stakeholders' Interpretations of Data for Equitable Computing Education
Benjamin Xie (he/they)
University of Washington
bxie@uw.edu

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.

Codeitz informed agency in online learning
Student Amp contextualizing student feedback
DIF contextualizing test bias w/ domain expertise

gallery
Supplementary Slides
Now that I’ve defined my framing of equity and minoritized groups, I need to acknowledge my own positionality. I’m part of dominant groups. I’m an able-bodied Asian man who learned computing at top tier universities; in practically any setting, I can say “I know stuff about computing” and people will believe me. So computing education is made for me!

So when I’m doing research to design for minoritized groups, I’m designing for someone else essentially. So I have to approach this work with humility, partner with colleagues with diverse expertise and lived experiences, and consult many funds of knowledge.

There are three commitments to my research:

1. Societal structures affect people differently. This follows a Foucault tradition of investigating power relationships at the margins, at the inflection point of normal and abnormality, to problematize what and whom we exclude and object to.

2. Data is imperfect and biased and even creates realities of its own.

3. People are more than the data that we often use to represent them. This idea follows in the tradition of critical data studies.

And with all this being said, I still believe that data in its imperfect and biased nature can help stakeholders take equity-oriented actions!
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.

<table>
<thead>
<tr>
<th>study</th>
<th>key finding</th>
<th>prior knowledge</th>
<th>perceptions of power relationships</th>
<th>cultural competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codeitz</td>
<td>learning not affected by agency</td>
<td>agency unfamiliar, deviated from expectations</td>
<td>(did not consider)</td>
<td></td>
</tr>
<tr>
<td>DIF/test bias</td>
<td>designers identified potential changes to curriculum, test design</td>
<td>drew upon knowledge of curriculum, test design</td>
<td>focused on what they could control</td>
<td>varied, but enabled broader consideration of bias</td>
</tr>
<tr>
<td>StudentAmp</td>
<td>teaching teams considered how challenges affected subgroups differently</td>
<td>drew upon prior experience taking and teaching course, at institution</td>
<td>focused on changes within course</td>
<td>prior training enabled deeper consideration of student identity</td>
</tr>
</tbody>
</table>
Interpretations and Uses of Data for Equity in Computing Education
Benjamin Xie (he/him)
University of Washington
bxie@uw.edu

We can enable informed, timely, and equitable action by designing interactions with data that enable stakeholders to connect their interpretations of data with their domain-expertise.

Codeitz
informed agency in online learning
agency must be an informed option

Student Amp
amplifying voices of marginalized groups
contextualizing feedback w/ identity provides benefits but also risks

DIF
contextualizing test bias w/ domain expertise
data on bias requires judgement of domain experts
Codeitz: informed agency in online learning
Agency must be an informed option

Student Amp: amplifying voices of marginalized groups
Contextualizing feedback w/ identity provides benefits but also risks

DIF: contextualizing test bias w/ domain expertise
Data on bias requires judgement of domain experts

<table>
<thead>
<tr>
<th>Codeitz</th>
<th>Student Amp</th>
<th>DIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>informed agency in online learning</td>
<td>amplifying voices of marginalized groups</td>
<td>contextualizing test bias w/ domain expertise</td>
</tr>
<tr>
<td>Agency must be an informed option</td>
<td>Contextualizing feedback w/ identity provides benefits but also risks</td>
<td>Data on bias requires judgement of domain experts</td>
</tr>
</tbody>
</table>

@benjixe
bxie@uw.edu
world view
all content always available (similar to MOOC)
instruction to read, exercises to practice
no world view
similar to basic ITS
system decides next exercise
IL: informed low-agency

On submit, updates skill bars and click “next” to try next exercise
RQ1 results: themes of challenges

- course-related
  - course structure
  - course content
  - remote learning

- external
  - other classes
  - extracurriculars
  - home & family
  - job
  - location
  - political
  - COVID-19

- individual
  - physical health
  - mental health
  - isolation
  - motivation
  - time management
RQ2 results: factors impacting sharing

- **privacy** of selves, classmates, others
- perceptions of what teaching team should know
- balancing vulnerability w/ need for help
identity impacts lived experience

students   teachers   designers

environment
DIF is often used by educational testing companies to ensure high stakes exams are fair. For example, say someone at ETS was creating questions for this upcoming years’ AP Computer Science Principles exam. They would likely use DIF techniques to identify questions that exhibited DIF and remove them because they may disadvantage certain groups (by gender or race for example).
Some potential future work includes engaging additional stakeholders, using human-centered AI techniques to enrich data with domain expertise, and monitoring the effect of interventions.
Now say we wanted to look at a group of students who reported as male and a group who reported as female. All students got questions A and C correct. We would expect them to get question B correct as well. And say all the students who reported as male do get B correct. But say we observe that most students who report as female get question B wrong. This is a toy example that demonstrates Differential Item Functioning (DIF), where students on similar knowledge levels but different groups (genders in this case) perform differently on an item, question B in this case. DIF is a technique to identify potential bias in test questions.
uniform DIF; reported females of all knowledge levels \( \theta \) have lower probability of getting items correct

female

male
agency: self-efficacy and information are critical

- **agency**: a learner can take actions that align with their learning-related goals (*Wardrip-Fruin et al. 2009*)
- **self-efficacy**: belief in ability to organize and execute course of action, process information (*Bandura 2001, 2006*)
- **information**: proximal action-related key to making decisions (preference construction, Bettman, Luce, & Payne 1998; Lichtenstein & Slovic 2006)
high-agency conditions completed more practice

num exercises completed

test scores

- UH
- IH
- IL
test scores: no difference across conditions

Potential explanations:

• most learners finished all exercises
• learners did not exercise agency
• assessment did not measure well

did not find diff in learning outcomes by condition
- no agency: used to following instructions at undergrad studies
See paper for more explanations and qualitative data about this
prior programming experience, greater self-efficacy predictive of higher test scores
Codeitz: navigation enables flexibility

low self-efficacy: follow recommendation

high self-efficacy: decide for themselves
importance of features

world view
“helpful to see how concepts fit together” (UH)

progress indicators
(all conditions found helpful)

exercise feedback
“hints and better feedback when you get an answer incorrect... would help me feel more confident” (UH)

recommendations
“jump around” (IL), “jump too far” (IH)

skill bars
“helped me know whether or not I should move on to the next topic” (IH)
participant feedback on importance, role of feature
- world view: high-agency only
- progress: useful across all conditions
- exercise: generally helpful, wanted more hints and feedback to fix mistakes

info based on system predictions (only for informed conditions)
- rec: least helpful of the features (paper)
- skill bars: move on or not
data: 15 feedback session across 5 courses
analysis: inductive thematic analysis with subsequent round of qualitative coding using themes from initial analysis

Started with 3 researchers (Alannah Olesson, Jayne Everson) collaboratively affinity diagrammed 100 random challenges to generate initial themes. We then collaboratively coded 5% of the data, discussed discrepancies, and iteratively refined the code set and definitions. From there, two researchers collaboratively coded 15% of the data, checked in with each other, then divided the remaining challenges between each other to code.
here's what the instructor would see after students share feedback using Student Amp:
see results in feedback session
challenges
demographics
disrupt score
label challenges
see how labels disproportionately affect certain groups
workload
BIPOC, work full time, transfer, moderate mental/social disability
RQ1 data: What did students share

data: 810 challenges from 604 students

"timezone / couldn't really access office hours sometimes because it's so late"
S-A-189, Asian woman, minor mental and physical disabilities, first-generation, non-English familial language

My father will be going out of the country next week on [date]. When he is usually home, he watches my sister when she is in class, and so now that he will be gone, I have to do that, which takes away half of my week.
S-D-36, BIPOC first-generation woman, works part-time, minor physical disability

I'm unsure of my ability to train my brain to think this way.
S-B-31 white woman, severe mental and minor physical disabilities, transfer student, 1st programming course

data: 15 feedback session across 5 courses
analysis: inductive thematic analysis with subsequent round of qualitative coding using themes from initial analysis

Started with 3 researchers (Alannah Olesson, Jayne Everson) collaboratively affinity diagrammed 100 random challenges to generate initial themes. We then collaboratively coded 5% of the data, discussed discrepancies, and iteratively refined the code set and definitions. From there, two researchers collaboratively coded 15% of the data, checked in with each other, then divided the remaining challenges between each other to code.