Understanding and Supporting Equity in Science Classrooms with Visual Learning Analytics: A Novel Approach Using Student Electronic Exit Tickets (SEETs)

by

ALI RAZA

B.S., University of Lahore, 2014

M.S., University of Management & Technology, 2016

M.S., University of Colorado Boulder, 2021

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Computer Science, Institute of Cognitive Science

2022

Committee Members:

Tamara Sumner, William R. Penuel (co-chairs)

Clayton H. Lewis

Danielle Szafir

June Gruber

Ali Raza (Ph.D., Computer & Cognitive Science [Department of Computer Science, Institute of Cognitive Science])

Understanding and Supporting Equity in Science Classrooms with Visual Learning Analytics: A Novel Approach Using Student Electronic Exit Tickets (SEETs)

Thesis directed by Professor Tamara Sumner, William R. Penuel

Abstract

This dissertation examined how a visual learning analytics system and subsequent teacher collaboration in a professional learning cycle can support improving the quality of students' experience in middle school science. It consists of three separate studies. In the first study, we followed an iterative co-design process over a year with seven middle school science teachers in gathering feedback on 30 visualizations of the student experience to identify critical visualizations for engaging teachers in reflecting on how equitable their students' experience was, to inform the design of a dashboard for teachers. In the second study, my colleagues and I investigated 42 middle school science teachers' use of the *Science SEET system* and corresponding visualizations to support equitable instruction over a two-month long longitudinal study. Findings show that only equity visualizations prompted teachers' reflections on diverse student experiences. However, despite the support equity visualizations provided for this core task, the teachers consistently ranked the "whole class" visualizations as both more usable and useful in their practice due to a variety of factors, ranging from their perceived usability of each visualization type, the teachers' own perspectives on equity, and the demographics of their particular classroom. In the third study, we investigated whether engaging teachers in two separate cohorts (n=24, n=25) in a novel disciplined inquiry cycle, the *Student Experience Improvement Cycle (SEIC*), could improve the quality of student experience, particularly for racially minoritized students and students of different genders. Findings show that SEIC supported improvements in the quality of students' experience for both teacher cohorts and, in three cases, differentially benefited Native Hawaiian/Pacific Islanders, Asian, and Black or African American students. We conclude that formative use of student experience data can improve the quality of student experience in classrooms but that achieving equity goals is more challenging. Based on the three studies, I conclude: *Equity-based visual learning analytics and teachers' collaboration in a disciplined inquiry cycle using evidence-based instructional strategies can support equity of participation in science classrooms*.

Key Words: equity analytics, visualization, students' experience, learning analytics, improvement science, science education, learning sciences, design-based research, teacher learning, data use for equity, professional development, teachers' collaboration.

Dedication

To all my Teachers, who made me see the world through their eyes

and

To Parveen and Samina for their kindness and love (grandmother and mother)

Acknowledgement

First, I express my deepest gratitude, love, and respect to my advisors: Tamara Sumner and William R. Penuel, for their amazing support, encouragement, and guidance throughout my graduate school experience; this thesis would not have been conceived, let alone written. I have learned so much from you in these years, both academic and lifelong lessons, especially in tackling the most pressing research problems with a smile and enthusiasm. You have shaped my thoughts profoundly, and thanks for showing infinite patience with me. Your discipline and love for the field have made me an improved scholar and human. Thanks for taking care of me and making it feel like home. If I were half a shadow of both of you, I would be a great teacher one day.

I am also forever indebted to my committee members: Clayton Lewis for trusting me with my research abilities when I was not at CU and encouragement as a young scholar in your class, Danielle Szafir for introducing me to Information Visualization and being an amazing motivation in the field to me, and June Gruber for being a fantastic mentor to me in these years, your kindness and concern for the students' good are exemplary. With the support, ideas, and guidance of all of you, this work has reached new heights.

Thanks to Jiawei Li and Ken Frank, for your amazing year long partnership in helping us with the analysis of the SEIC cycle and I learned much from you about HLM, although there is still so much to learn. I also want to sincerely thank Eddie Dombrower for working with me for more than a year on the Science SEET system, and whenever I was stuck for day(s) on some error(s), you would show up with a smile and support me. Finally, I want to remember and thank Michael Eisenberg (late) for having a significant influence on me in graduate school with his class

(Technology and Young); as I recall, practically at every single round table in class, there was a student(s) who was audio recording his voice in those classes.

I also want to thank Spencer Foundation and Bill & Melinda Gates Foundation for providing valuable funding that resulted in the work we share through the three studies in this dissertation.

I am very grateful to all the participating teachers in the co-design process, and in two cohorts of the SEIC cycle for your time, commitment to the work and the amazing energy. I am also very thankful to all the students' who allowed us to learn about their experience in the science classrooms.

During graduate school, I was fortunate to be part of outstanding research groups that shaped different aspects of this dissertation. InquiryHub (iHub), Advancing Coherent and Equitable Systems of Science Education (ACESSE). I am also extremely grateful to all the lovely friends I made in the Sumner Lab (Rick Parker, Jennifer Jacobs, Ahmed Mohamed Fahmy, Srinjita Bhaduri, Michael Schneider, Varsha Koushik, Abhijit Suresh, David Quigley, Alex Gendreau, Quentin Biddy, Jay Luther, Jeff Bush, and Emily Jensen) and Positive Emotion and Psychopathology (PEP) lab (Cindy Villanueva, Stevi Ibonie, Ansel and Silvan Gruber), you guys pushed me to think better with your feedback. Thanks to Riyaz Bhat for your kindness, for being my best roommate at the start of graduate school, and for introducing me to U.S. culture. Thanks to Kerri Wingert for all the laughs we and our kiddos exchanged during these years. Thanks to Fahad and Ayesha for making our weekend meal lit up with joy and being amazing friends.

I want to thank my high school level home teacher, Sir Iftihikar Anjum Ghouri, for teaching me Science and Math. Your persistence and hope for all students in less privileged areas still greatly inspire me. I also want to thank Dr. Shoaib Farooq for introducing me to the world of research and specifically working with me in Computer Science Education research during my master's degree back home and motivating me to think expansive.

I want to thank my grandmother (Parveen Akhtar), grandfather (Rasheed Ahmed), and mother (Samina Rasheed) for raising me and everybody in the family (Ejaz Rasheed, Saima Rasheed, Sumera Rasheed, Tehmina Ejaz, Khudaija Ejaz, Zainab Ejaz, and Hassan Raza) for giving me space in their hearts. Zainab, it is always so happening talking to you, thanks for laughs, encouragement and we are all so proud of you, can't wait to see what's next for you. I also want to thank my in-laws' family, Dr. Rubina Karim, Shaheena Raza, and Raza Raheem, for their amazing support in this journey. Thanks to my childhood friend, Rao Abdul Moeed, for sharing our dreams and emotions in all these years.

Last but definitely not least, I want to thank Ifrah Ali Raza for being the best partner I can imagine. Thank you for showing all the patience with me, infinte support, and always loving me unconditionally, and listening to all my dissertation talking in these years. Aaira Ali Raza and Zoha Ali Raza, thank you for being in our life, hearing your voice every morning makes me so happy, and I can't wait to see you grow and the goodness you will bring to this world.

I have missed many people on these acknowledgement pages, who helped and supported me in my educational endeavors from Pakistan and the U.S. I am deeply thankful to all of you.

CONTENTS

TABLES

FIGURES

CHAPTER I: INTRODUCTION

Overview of Dissertation and Significance

Students can have inequitable experiences based on race and gender during knowledge-building activities in science classrooms. For example, Wright and colleagues (2018) give an account of a group of girls in an engineering class who decide not to engage in arguing for alternative designs for their project, for fear that their "arguing" would be perceived by the teacher as disruptive. These girls were well aware of racialized and gendered scripts that could be consequential for them, and so they resisted participating in ways that were intended. For students like these impacted by inequitable treatment, bias, and stereotyping have often limited their access to learning opportunities, classroom participation, and career pathways in STEM or science fields (Greider et al., 2019; Master & Meltzoff, 2020; Gruber et al., 2020; Shah et al., 2020). They lead to questions from students as to whether they belong in science or if science is for students like them (Brown, 2019).

There is an increasing focus on supporting equitable learning experiences for Black, Indigenous, People of Color (BIPOC), and underrepresented female and non-binary students and their communities (NRC, 2012). One approach to supporting an equitable learning environment is to gather and reflect on student experience data and adjust instruction to ensure different groups of students' experience is equitable, in terms of how they participate. Recently, an approach called 'Equity analytics' (Reinholz & Shah, 2018) is proposed to quantify classroom participation across social markers such as gender, race, and other social markers to better understand students' participation in knowledge-building activities in STEM classrooms. This approach has been shown to help teachers become more reflective about the degree to which their classroom instruction is equitable in mathematics classrooms (Reinholz & Shah, 2021).

In this dissertation, I worked with colleagues to support the equity of participation in science classrooms by using a visual learning analytics system; the *Science Student Electronic Exit Ticket* (Science SEET) system, and responding with research-based instructional strategies in a disciplined inquiry cycle; *Student Experience Improvement Cycle* (SEIC). The SEIC is based on a Plan-Do-Study-Act cycle (Bryk et al., 2015) that engages researchers and teachers in adapting research-based strategies for promoting more equitable small-group and whole-class knowledge-building activities. As part of the SEIC cycle, two cohorts of participating teachers investigated research-based strategies for improving equity of participation in science classrooms, using student experience data they gather from short surveys in their classrooms to assess whether the strategy helped improve the overall student experience as well as differences in experience associated with students' race and gender. The SEIC includes: (1) the use of a web-based SEET system that supports educators in collecting, analyzing, and automatically visualizing student experience data from brief surveys administered at the end of class (as a whole, and broken down by race and gender), (2) embedded research-based summaries, ideas, and strategies that can be drawn upon in planning instructional responses based on experience data, and (3) collaborative learning processes for researchers and teachers. The Figure 1 below provides a brief overview of the activities teachers and researchers engage in four meetings during the two months, including the use of the SEET system.

The SEET system provides teachers with several ways of visualizing their student data, including three 'equity visualizations' that have been carefully designed to support teachers to notice differences in student participation across race and gender, and three 'whole class visualizations' that depict student experience data for the class as a whole (Table 2). Both types of visualizations emerged from a year-long, iterative codesign process with middle school science teachers that are described in chapter 2 (Raza et al., 2021). These visualizations render three aspects of student experience data collected with surveys: coherence, relevance, and contribution, which are important to a particular vision for science teaching and learning as articulated in *A Framework for K-12 Science Education* (NRC, 2012). Coherence refers to the students' perspectives of the progressions of their learning experiences, and whether these progressions were driven by student questions, ideas, and investigations (Reiser, Novak, & McGill, 2017; Reiser et al., 2021). The relevance construct is rooted in expectancy-value theory (Eccles, 1983), which highlights the importance of enabling students to make connections between their interests and identities, and classroom activities. When students perceive instruction to be relevant to them, it can lead to motivation, engagement, persistence, and self-regulation in learning (NASEM, 2018). We chose 'contribution' as our third construct as students' contributions in science classrooms as partners in knowledge building are considered critical in contemporary science reform (Schwarz, Passmore, & Reiser, 2017), as classrooms shift from a model where the teacher is the holder of knowledge to one in which students are epistemic agents, i.e., "individuals or groups who take, or are granted, responsibility for shaping the knowledge and practice of a community" (Stroupe, 2014, p. 488).

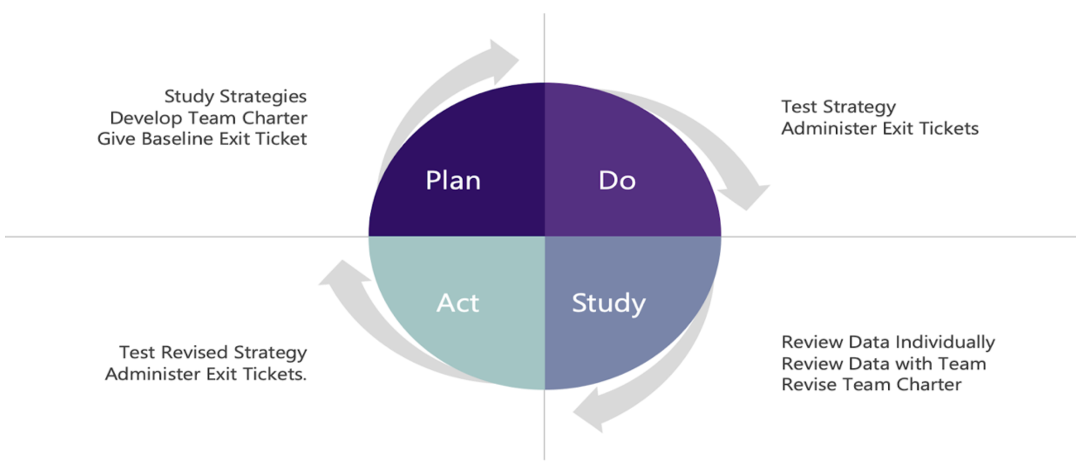


Figure 1: The SEIC cycle for supporting equity of participation in science classrooms

This dissertation aims to demonstrate that improvements to the quality of students' learning experiences are possible, including for students from nondominant groups and communities. With the use of novel co-designed visualizations in a learning analytics system that effectively prompted teachers' reflections on inequitable patterns of students' classroom experience and engaging in a disciplined inquiry cycle using research-based instructional strategies can positively impact the underrepresented students' experience in science education.

Research Context

The Science SEET system is motivated by and builds on prior work from the inquiryHub (iHUB)[1] research-practice partnership, where the SEET survey was initially developed to help the partnership understand variation in classroom implementation of the iHUB curriculum (Penuel & Watkins, 2019). The team was able to show that responses to survey items correlated with student learning measures (Penuel et al., 2018), and also that in fact there was wide variation across classrooms in student experience (Penuel et al., 2016). Until the Science SEET system was developed, the team relied on paper-based and Google forms to collect student experience data, which was a tedious process for practitioners to use to be able to interpret data to achieve any actionable insights. Thus, our team sought to design and develop computational infrastructure to gather and manage these data effectively and to display student experience data in ways that could support teachers in noticing patterns of equity/inequity in classrooms.

The three studies in this dissertation emerged from two different projects, 1) Students' Experience as Formative Assessment (SEFA), 2) Advancing Coherent

[1] iHub is a mature research-practice partnership between University of Colorado Boulder and Denver Public Schools. With significant members consisting of teachers, district administrators, and university researchers. One of the products of this partnership is an inquiry driven high school science curriculum based on Next Generation Science Standards. For more details, see: https://www.colorado.edu/program/inquiryhub/

and Equitable Systems of Science Education (ACESSE). The former focused on exploring the use of student experience data as a formative assessment and on designing and developing the Science SEET system for building this capacity. Study 1 is associated with the SEFA; we worked with the seven middle school science teachers in Denver Public schools as part of the research practice partnership to co-design visualizations of the SEET system and then later implement the SEET system. The second project built on the first, but was conducted within a different partnership between science leaders of the Council of State Science Supervisors and researchers at the University of Washington Seattle and University of Colorado Boulder (Penuel et al., 2018). In the project, our team worked with middle school teachers from across the country to use the Science SEET system, designing the SEIC cycle to engage two different middle school science teacher cohorts (n=25, n=24) each over two months to learn about their SEET system use and study how teachers' collaboratively use evidence-based instructional strategies in small groups of 3-5 to support equity of participation.

Research Questions and Studies Completed

This dissertation investigates questions in three studies to learn about the affordances of co-designing visualizations with teachers, the use of a learning analytics tool, the ScienceSEET system, to support teacher reflection of equity of participation in their classrooms, and to document evidence related to whether participation in the SEIC was associated with improvements to quality and equity of participation in science classrooms.

Research Question 1: What visualizations during co-design support teachers in noticing patterns and trends related to equity in students' experiences in science classrooms? (Completed, Full paper in chapter 2, published at *International Conference of Learning Sciences* (ICLS))

Study 1: Following an interactive two-part co-design process over a year with seven middle school science teachers in gathering feedback on 30 visualizations of the student experience, we explored how different types of visualization (such as bar charts, line charts, bubble charts, choropleth heatmap, etc.) help teachers to reflect on classroom equity, and which visualizations are preferred. Findings showed that it is imperative for researchers and system designers to have an adaptive view when designing equity-oriented feedback dashboards and to deeply attend to teachers' perspectives and guidance in order to ultimately support equity in the classroom in meaningful and relevant ways.

Research Question 2: How can different visualizations of an equity analytics tool support middle school science teachers in reflecting and using the experiences of different groups of students in their classrooms? (Full paper in chapter 3, under review at *Learning Analytics & Knowledge conference*, (LAK, 2023))

Study 2: This two-month-long longitudinal, mixed-methods study examines how forty-two middle school science teachers notice differences in students' classroom experiences when presented with equity visualizations disaggregating data by gender and/or race, and do they prefer visualizations that break down student experience data by race or gender to the whole class ones that do not. Findings show that only equity visualizations prompted teachers' reflections on diverse student experiences. However, despite the support equity visualizations provided for this core task, the teachers consistently ranked the whole class visualizations as both more usable and useful in their practice due to a variety of factors, ranging from their perceived usability of each visualization type, the teachers' own perspectives on equity, and the demographics of their particular classroom.

Research Question 3: Can a professional learning cycle support teachers to use student experience data formatively collaboratively as part of the cycle of inquiry to increase equity with respect to students' experiences of and contributions to knowledge-building activities in the classroom? (Full paper in chapter 4, to be submitted at *Journal of Research in Science Teaching*)

Study 3: We investigated whether engaging teachers in two separate cohorts (n=24, n=25) in a cycle of collaborative inquiry related to strategies for promoting more equitable participation in science could improve the quality of student experience, particularly for racially minoritized students and students of different genders. Findings show that SEIC supported improvements in the quality of students' experience for both teacher cohorts and in three cases, differentially benefited Native Hawaiian/Pacific Islanders, Asian, and Black or African American students. We conclude that formative use of student experience data can improve the quality of student experience in classrooms but that achieving equity goals is more challenging and requires multiple iterations of the cycle.

Dissertation Outcomes

This dissertation yields numerous outcomes, including theoretical contributions, system-building contributions, and associated professional learning resources.

- A robust Science SEET system, iteratively refined through teacher co-design and ongoing use. This system offers an exemplar in the equity analytics field. This research contributes to our understanding of how teachers can use technology to support effective decision-making for supporting equity of participation in STEM classrooms.
- Understanding teachers' use of visualizations of student experience related to equity from their classrooms. It helps us learn what factors might impede the

teachers' use of equity visualizations and how to support their sensemaking with respect to differences in experience by race/ethnicity and gender and to use this data effectively to support equitable instruction.

- Demonstrate the types of professional learning and collaboration support in the SEIC cycle that can enhance teachers' effective use of the system. This helps us to understand how collaborative inquiry cycles can support teachers to effectively integrate the SEET system into their practice.
- Initial, evidence-based conjectures about the conditions under which equity of student experience in classrooms can improve. It contributes to our knowledge of how an explicit equity stance: seeking to promote equality of experience and, where possible, stronger support for students most excluded from STEM, can contribute to making instruction more equitable.

Outline of the Dissertation

In chapter 2, the study investigates the affordances of the co-design visualizations of the student experience that ultimately are implemented in the Science SEET system. In chapter 3, the study investigates what differences do teachers' remark when reflecting with different visualizations of the SEET system and which one's teachers prefer in their use for supporting equitable instruction. In chapter 4, the study explains how collaboratively working in the Student Experience Improvement Cycle (SEIC) cycle supported teachers to positively impact historically underrepresented students' experience in STEM classrooms. Finally, in chapter 5, discussion and conclusion, I summarize the findings from each study and then engage in a discussion building on the three studies and future implications of the work.

CHAPTER II: “Making it Culturally Relevant”: A Visual Learning Analytics System Supporting Teachers to Reflect on Classroom Equity

Ali Raza, William R. Penuel, Anna-Ruth Allen, Tamara Sumner, Jennifer K. Jacobs
University of Colorado Boulder

Abstract

Experience is an important dimension of learning (Roth & Jornet, 2014). Research has drawn attention to the importance of equitable learning environments where all students experience opportunities to contribute to knowledge-building activities. Working in partnership with five middle school science teachers and two instructional coaches, we have developed a visual learning analytics system that aims to support teachers in reflecting on variability in students’ classroom experiences linked to race and gender. This system, the Student Electronic Exit Ticket (SEET), captures and visualizes student experience data related to three constructs that prior research indicates are reflective of equitable classroom experiences: coherence, relevance, and contribution. We share findings from a qualitative study of how the SEET system visualizations supported teachers and coaches to recognize and respond to classroom inequities. Further, we discuss our design process with teachers and how this partnership influenced the visual feedback display.

Introduction

It is critical to make science meaningful to students from a wide variety of backgrounds, identities, languages, and cultures. Equitable practices in science classrooms are supported when we are able to make science culturally relevant to all students. Students can have inequitable learning experiences when teaching and subsequent actions are idealized in potential harmful ways based on learner attributes, such as geography, zip code, language, religion, gender, race, ethnicity, or socio-economic status (Brown, 2019). For instance, a student's way of speaking or their choice of clothes can deem them as not intelligent to the teacher, and these perspectives on ability can be rooted in biases around culture and race. This has been called the "Black Tax" in science classrooms, which refers to the "additional hurdle or cost faced by many students of color" (Brown). This black tax narrative is not only restricted to Black communities, instead it frames the narrative of many non-white populations who experience challenges (Brown, p.13).

In order to disrupt this narrative and create equitable learning environments, teachers need ways to more deeply recognize, understand, and engage with their students' experiences of the classroom. Attending to students' experiences involves integrating the practical, intellectual, and affective dimensions of classroom life (Roth & Jornet, 2014). Understanding students' experience in the classroom can help science teachers alter their instruction in ways that make it more culturally sustainable for a diverse student body, across gender and race (Paris & Alim, 2017). Penuel and Shepard (2016) note that helping teachers to understand the socio-cognitive dimensions of classroom experience (attending to the social nature of cognition) and the socio-cultural dimensions (attending to the value of participation in disciplinary ways of knowing and doing) of their students' experiences can support efforts to promote equity in classrooms.

Taking a student-centered approach and gathering reliable information about students' experiences in the classroom sheds light on the learning community and processes from students' perspective (Penuel & Watkins, 2019). Drawing on recent efforts to produce 'Equity Analytics' as a quantitative approach to view patterns of equity and inequity in classroom participation across gender and race (Reinholz & Shah, 2018), we have developed an innovative system of surveys and visualizations that captures and visualizes student experience data revolving around three constructs: coherence, relevance, and contribution. The Student

Electronic Exit Ticket (SEET) system includes three major components: multiple versions of a student survey that can be administered anonymously in classrooms, a data management component, and a corresponding analytical system for teachers that visualizes these survey responses and disaggregates them by gender and race. In this study, we examine how different visualizations of student experience data during design process of the tool can support teachers to understand and reflect on the experiences of diverse students in their classrooms. We also report on how adopting a user-centered design process helped in selecting and adapting the user interface of the SEET system.

In our research, we worked closely with five middle school science teachers and two instructional coaches to design the visualizations and understand affordances offered by different representations for identifying patterns of equity and inequity of experience in science classrooms. Our ultimate goal is to develop analytic and instructional routines that help teachers to interpret these data and use them to reflect on their classroom instruction. The following research questions guided this paper: 1. How do different data visualizations create different opportunities for teacher's sense-making? 2. In what ways does visualizing students' experience data promote teachers' reflections on equitably supporting students' needs?

Conceptual Framework

We argue that supporting equity involves understanding learners' experiences based on three constructs: coherence, relevance, and contribution (Reiser et al., 2017, NASEM, 2018, Miller et al., 2018). To this end, we have created an experience based formative assessment premised on these constructs called the Student Electronic Exit (SEET). SEET data provide targeted information about learners' experiences within a particular academic unit and classroom. Each construct comprises a unique set of questions. SEET questions related to coherence ask students whether they understand how current classroom activities contribute to the purpose of the larger investigations in which they are engaged. Coherent learning experiences appear connected from the students' perspective, where the progression of learning experiences is driven by student questions, ideas, and investigations (Reiser et al., 2017). Questions related to relevance ask students to consider the degree to which lessons matter to the students themselves, to the class, and to the larger community (Penuel et al., 2018; NASEM, 2018). For contribution, SEET questions ask students whether they shared their ideas in a group discussion, heard ideas shared by others, and whether others' ideas impacted their thinking. The aim of using the SEET assessment in the classroom is not to judge teachers or identify students' understanding of disciplinary content. Rather, the SEET assessment is intended to help create an environment for improving teacher instruction and diminishing classroom inequity.

Overview of the Science SEET system

The SEET system uses experience sampling to gather information about students' classroom experiences. Experience sampling is a technique for gathering data about how people feel and think during random intervals in their lives (Larson & Csikszentmihalyi, 2014). In our context, the teacher decides when to sample students' classroom experiences. The SEET system uses a thirteen-item survey to gather information from students on their perceptions of coherence, relevance, and contribution for that day's classroom experiences. This survey is designed to take only a few minutes for students to complete at the end of a class session. It is important to note that the SEET system does not collect names of students, provide individual-level feedback to teachers, nor does it display disaggregated data across gender and race if there are fewer than three students belonging to a particular gender or race to protect the anonymity of the students. Key steps in the workflow for teachers and students are: (1) The teacher administers a survey that asks students to reflect on today's classroom experiences, (2) Each student in the class receives the survey and reports on their experience, (3) Student responses from the entire class are combined and displayed to the teacher through a visual analytics dashboard based on the three focal constructs, (4) Teachers reflect on the student experiences' with visualizations.

Data and Study Methods

We recruited five middle school science educators (two identified as female and three as male) and two science instructional coaches (one female and one male) from a large, urban school district in the Midwest to participate. The materials for this study included 30 different visualizations across three constructs. Round one included visualizations from a made-up dataset created by the first author that had significant differences in student experience across gender and race. For round two, we drafted visualizations using actual classroom data collected from a middle school science teacher's classroom in the same school district. This teacher was not a study participant.

In Round One, three science teachers (one female and two male) and two science instructional coaches (male and female) were asked individually and in pairs to reflect on selected visualizations from a pool of 14 different visualizations. Each session took 30-35 minutes. Teachers were shown a visualization and asked to think aloud as they tried to interpret its meaning and its implications for teaching. Think aloud protocols are a technique for understanding the cognitive processes of the participant when a stimulus is introduced during decision making (Ericcson & Simon, 1993). After each think-aloud, teachers completed a short interview about each visualization. During the interviews, we probed teachers about their thinking, asking them “What are you wondering about, assuming this is a real classroom? What is the need in your classroom that this visualization points to, if this were your classroom’s data?”. After seeing all the visualizations, we asked teachers to tell us which ones they believed to be most useful for helping them monitor their students’ classroom experiences.

For Round Two, the same investigative process was followed with two science teachers (one female, one male) and the number of visualizations was increased to 30 (the 14 from Round One and 16 additional visualizations). New visualizations were created to address teacher feedback from Round One.

To gain deeper insights into how specific visualizations supported teachers’ reflection processes, we analyzed our qualitative data with an inductive lens using a grounded theory approach to identify themes (Charmaz, 2014). The first author went through all the transcripts line by line to develop a preliminary code book. This preliminary codebook was developed into a mutually exhaustive and exclusive code book as shown in Table 1, after two authors met multiple times to reach consensus. The high-level codes included “Data reflection”, “Interpretability and Recommendations”. Within each high-level code there are multiple sub-codes. The two coders established high levels of interrater agreement calculated using Cohen’s Kappa; κ= 0.88 for Data Reflections and κ= 0.89 for Interpretability and Recommendations.

Table 1: Coding teachers’ responses to the data visualizations

High level codes	Sub-codes	Examples
Data Reflection	• None: Teacher finds nothing • Find: Teacher finds a particular data pattern. • Inquire: Teachers reflect on student experience data possible relationships between constructs and findings.	• Find: “If you answered questions last, did you share any ideas out loud today. That is really striking to me.” • Inquire: “I am curious about these don't know categories, what commonalities are there between those statements?”

	● Instructional need or aim: Teacher thinks on what can be the next steps. ● Pointing towards equity: Teacher talks about students from non-dominant groups and BIPOC, and/or when teacher talks about "equity" or "inequity" across gender or race.	● Instructional need or aim: "So, I would think that's like something on me to make sure I'm tying it back to like our bigger picture." ● Pointing towards equity: "Did anyone else share? 85% say others shared? To me, this kind of points to a sense of inequity."
Interpr-etability and recommendations	● None: Teacher doesn't interpret and provide recommendations ● Hard Identification: Hard to interpret the visualization ● Easy Identification: Easy to interpret the visualization ● Attraction or compelling point: Teacher reasoning about the compelling and attractiveness in visualization ● Anytime Principle: Teacher has some anytime principle in mind on how to read visualization. Sense-making process tied up with the anytime principle. (Russel, Stefik, Pirolli, &, Card, 1993) ● Design Suggestion: Teacher provide design suggestion for interpretation of the visualization ● Construct Suggestion: Teacher provide suggestion related to constructs of the survey item SEET	● Hard Identification: I'm wondering about if I think the volume bubbles are kind of a neat idea for an easy glance piece. But I'm also maybe struggling to find how helpful they are. ● Easy Identification: "One of the things just like overall style, I like looking at the colors, color gradation a lot better." ● Attraction or compelling point: "It's a lot better, a lot easier to figure out what's being visualized. Than the size" or "I like this, I think the most because it gives me data per population" ● Anytime Principle: "But it took like, my eyes kept fluttering back and forth, and back and forth, and back and forth. And then I was just jotting. We've got tables, bar graphs, what do I need to look at first? So finally, I just started looking at this table." ● Design Suggestion: "I would like them portioned out per population. so, you could quickly see the smallest population or the largest percentage" ● Construct Suggestion: "Knowing my students, that's the comfort area. If I don't understand that question, or I don't know it, I'm gonna choose."

Space left blank intentionally for start of new section

Results and Discussion

During round I, the first three teachers selected visualizations: Horizontal Stacked Bar Chart, Line Chart, as these supported them in the sense-making of data. Both coaches selected: Choropleth Heat map, Bubble chart, and Line chart, provided the most appropriate sense-making to them. From round II with two science teachers they selected visualizations: Horizontal Bar Chart, Choropleth Heat map, Line Chart. We finalized the 'Horizontal Bar Chart with '%yes' only', 'Choropleth Heat Map with disaggregation of gender and race', and 'Line Chart' into the visual feedback displays of SEET shown below created from a made-up dataset.

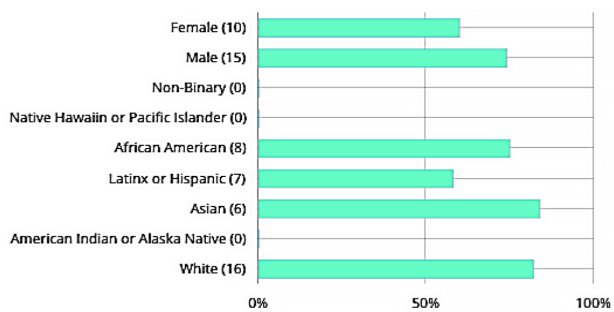


Figure 2: Bar chart by gender and race

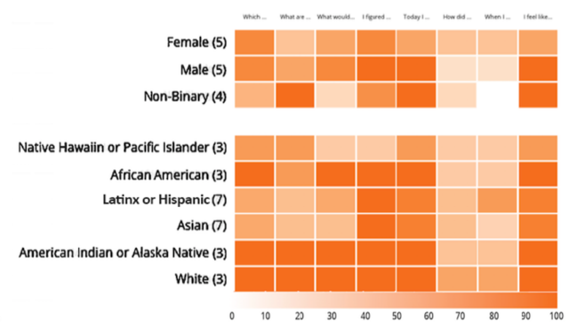


Figure 3: Choropleth Heatmap by gender and race

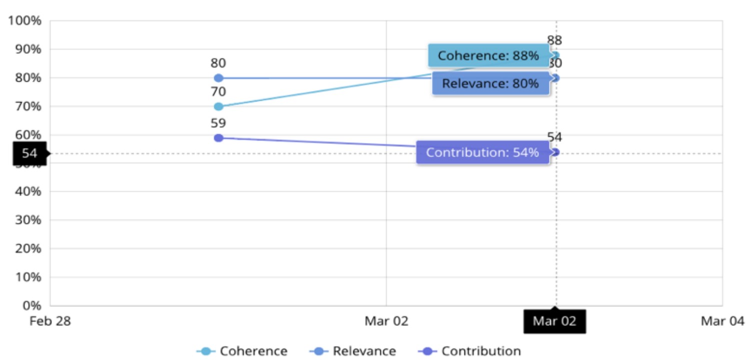


Figure 4: Over time by constructs

We found that the side-by-side bar charts facilitate easy comparisons, when they show a simple percentage of "yeses". We implemented small multiples of bar charts (Figure 2), adjacent to each other as suggested in prior work by Heer, Bostock & Ogievetsky (2010). The Choropleth Heat Map in Figure 3 displays data disaggregated by gender and race, the first three rows at the top display different genders, and then there is intentional space to start with different race categories. These categories are based on the U.S. census. Each column at the top belongs to lesson or measures (coherence, relevance, and contribution) related questions. Color facilitates thinking about the dimensions of experience that are being measured. Although the color is less precise as compared to position based visual representations such as scatterplots, bar graphs, and line graphs, they provide a significant 'big picture' information to the users (Albers et al., 2014). The line chart in Figure 4 is one type of implementation of its three variants: i) line chart visualized by correct average of SEET measures over time (see Figure 4) ii) line chart visualized by gender and race data over time for a single classroom period iii) line chart visualized by each question over time for a single classroom period. All these visualizations and variants are implemented in the SEET system.

Now we present data from teachers and instructional coaches for both rounds on how their interaction with different visualizations supported several reflections on equity and instruction prompted different reflections on student experience and how it helped in shaping the design of the user interface of the SEET system.

Data Reflection

Analysis of our qualitative data suggests that these selected visualizations in Figure 2, 3, 4 can be effective in supporting teachers to reflect on their students' classroom experiences. We now present a few cases where teachers expressed their perceptions related to student experience.

Mr. Alan (Science Teacher), all names are pseudonyms, when asked on what they see looking at data patterns in the horizontal disaggregated bar chart visualization (Figure 2), he responded:

> So, within the kind of the first boxes up here, there is a group of students who seem to be making more sense of the driving question and how it ties into less, and then other students. And it's looking like it's just the white students better following along better with the driving question board and how it ties into what they did in class and the unit, things like that. But then when we transition down to the other questions, matters to the people in the community and things like that. I’m not really seeing much there in terms of some overall takeaways, except for maybe the Black students. I don’t know how many are in there, though. Like, is that 18%? That one kid or total of five? Okay, there? Yes, I see it now. Total five kids. Okay. So, one of those five kids found that it tied back to their community and things like that, but the four of the five basically didn't notice that. (Mr. Alan, round II)

Mr. Alan in his reflection on what is the instructional need based on this classroom, responded:

> The instructional need would probably be making this culturally relevant to all students in the classroom, making sure that all students do find a connection to it, whatever they’re learning, because it’s clear that all students aren’t finding that connection from the lesson to what’s important to them or their community, only some students are, and it’s

> solely dependent on cultural background or ethnicity. (Mr. Alan, round II)

Mr. Alan showed concern about the Black students who didn't have the same percentage of "yeses" as the other students in the classroom. He tied the classroom experience to the classroom environment: saying it was not conducive to all the students from different backgrounds and cultures. In most parts of the interview Mr. Alan was concerned about the instructional connection to minoritized students such as Asian students and how to make it a more positive environment by making lessons coherent and equitable.

Similarly, when Mr. Meer and Mr. Tim (Science Teachers) are asked what they notice from the data visualization showing change over time, how they might use this data formatively for thinking or reflecting on things in classrooms, Meer responded:

> It's interesting that overall, we see it trending upwards. Like, obviously, I feel like that's each piece of the lesson building on itself, but super big drop off for lesson 3D for males and yeah, a big spike there for females so again, there had to be some kind of a gender disconnect when we get to lesson 3D. Yeah, lesson 3D obviously was off. So, it'd be a good starting point to look there. How can we make lesson 3D more equitable? I think it would definitely give me some empathy; especially like you said for here in coherence for females. (Meer, Round I)

While Mr. Meer pointing towards disaggregated bar chart question on coherence, he reflected:

> I would definitely drive my instruction to help to target this group, clearly. This was my classes with my data, I did not do my job enriching these female students. So, that would really cause me to go back and reflect and hopefully make some changes. This piece of data allows me to just directly ask them, what did you need that I didn't give you? (Mr. Meer, Round I)

Tim and Meer, both noticed the experience of the female students as not to the level of male students in the classroom. Meer was concerned about what he didn't do to make the classroom experience better for the female students and how he would inquire with students to make a better classroom connection across the parts of the unit to make an equitable environment for them. Mr. Tim and Mr. Meer both agreed that over time data can be helpful for them in identifying connections across the units and also disaggregated bar charts would help them target more specifically each classroom for coherence, relevance and contribution.

Mr. John and Ms. Becky (Instructional coaches), when asked about what's the instructional need based on the bubble chart representation, Mr. John pointed to inequity of experience related to the relevance for the students.

> Deeper piece and harder the relevance question six, there's an imbalance for who this seems to apply to. There's definitely a certain level of students who aren't feeling like this, this describes them. So that's like a larger piece. Like, which curriculum are we using? How do we want to make adaptations larger across our district. And then thinking about the training for those pieces. If we had the time to do professional learning. What implementations will we do? (Mr. John, Round I)

Mr. John's explanation pointed to the need for more relevant instruction for improving student experience on this construct. He indicated that few groups of students are not finding relevance with the lesson and how he can scaffold their experience with more relevant curricula that connects with their classroom experiences. He also wondered how professional learning and their implementations might help teachers for enhancing experience related to relevance. Listening to Mr. John's conversation, Ms. Becky, talked about the result of those interventions and how it can impact the classroom.

> Absolutely. And I think just techniques to incorporate some of these other pieces to help teachers make sure that all students are participating in, are able to connect to whatever it is that's going on. And also making it relevant too like, are we talking about an issue that the students can connect to that has a part of their normal life? (Ms. Becky, Round I)

Ms. Becky stressed on the importance of making sure that students contribute and make lessons relevant to their lives, and how instructional coaches can help teachers in facilitating these roles. Ms. Becky also narrated how student experience can be an opportunity for teachers to learn more and improve classroom environment while interacting with the Choropleth Heat map visualization.

> I think this is a really cool opportunity for teachers to gauge their classroom and their students. Like you can see there's a relatively high participation going on in the class verbally, but then there might be some understanding issues that are happening in kid's heads. So, the

> teachers. But I think this is a really informative opportunity for teachers to say, okay, we might need to work on this a bit more. Or how do I encourage my kids to speak up. I think it's informative for the teachers themselves. (Ms. Becky, Round I)

Ms. Becky pointed out how this representation is providing an opportunity for teachers to hear and know students' inner voices that can totally be different from their behavior in the classroom and be useful to teachers. As interviewer asked Ms. Becky and Mr. John on what instructional need this points to as your role as district leaders and instructional coaches, Becky responded:

> A lot of teaching strategies, you know, you could do a professional learning experience based on each one of these questions. And focus just on pulling that out for the teachers. Have you really included this method of instruction? or this piece of instruction. (Ms. Beck, Round I)

Ms. Becky and Mr. John both directed on understanding student experience data with different representations as opportunities for teachers to connect with their students. And how they can use targeted strategies with the teachers to meet the instructional needs of diverse and multicultural students. They noted that the engagement level of students tied with the success of students.

In summary, teachers we worked with cared about student experience related to coherence, relevance and contribution as much as they do about the content knowledge. Teachers know experience to be an integral part of academic success for students. Visualizations that disaggregated data by gender and race help them identify issues of equity in all instances, at least this was the case in our study. We found that these visualizations supported teachers to inquire about the experience of students and consider implications for instruction. They showed care for the experience of racially minoritized students and the experience of girls, and they valued student experience data that made this visible. However, how teachers can make use of this experience data related to coherence, relevance and contribution to enrich classrooms and student lives remains an underexplored and rich phenomenon. Visual learning analytic tools can be leveraged for the sake of supporting teachers' engagement with students' experience in the service of creating equitable and just learning environments.

Interpretability and Recommendations

Teachers' reflection on the interpretability of visualizations showed us where their interests lie and informs user-centered design of the SEET system. Below we discuss two examples of teacher noticing and suggestions that helped in developing the user interface of the system.

Example No 1: Horizontally categorizing bar chart by gender and Race

When the interviewer posted a data display containing disaggregated vertical and horizontal stacked bar charts to Ms. Yarosh (Science Teacher) to elicit her thinking and understanding would she prefer a horizontal or vertical bar chart, she responded:

> And I'm wondering if they're bars stacked on top of each other. So that takes quite a bit of thinking to unpack what's happening there? Um in question three. Well, I would say that these could all be bars but proportioned out between lesson three A. So, a bar that is proportionally representative of 20%, and then one that would show stacks on stacks on top of each other, not on top of each other like this but next to each other. So, you get that quick visual like, woah 63%. A quick visual for

> each lesson, right? So, I think either way, I would like them portioned out per population. So, you could quickly see the smallest population or the largest percentage. (Ms. Yarosh, Round I)

Ms. Yarosh initially expressed her thoughts that vertically stacked bars categorized by each lesson are hard for her to interpret the differences across race levels. She suggested that rather than stacking race categories on top of each other vertically on a bar, she would place them next to each other for a quick comparison across student populations to quickly compare. This conversation and suggestion of the teacher led to the design of a horizontal disaggregated bar chart based on categories for gender and race, see Figure 2.

Example No 2: Showing only "yeses" or correct responses facilitates comparisons.

A constructive suggestion by Ms. Yarosh shaped the design of visualization, as students were answering multiple choice questions that had three response options ('Yes', 'I don't know', and 'No'). During the first round, we displayed all three responses, conjecturing that it would help teachers identify patterns across the response types (Figure 5(A)). The yellow color represents the percentage of the class who said 'Yeses', Cyan represents 'I don't know', and Blue represents 'No', to a survey question based on a particular construct. Contrary to our expectations, teachers preferred visualizing only the 'yes' responses.

> So, I guess for me, this neutral choice doesn't necessarily give me the information I'm seeking. I would prefer it matters or it doesn't. (Ms. Yarosh, Round I)

Analyses of additional qualitative data suggested that displaying only the "yes" responses helped teachers to compare patterns across the three constructs; many instances teachers also showed that approach helped spark discussions about the low percentage of “yeses” answers from the students, as they perceived it a simpler solution for cross comparison across questions. This also facilitated cross comparison between constructs; coherence, relevance, and contribution by pointing teachers’ attention only to one value. The resulting visualization based on teacher feedback was a horizontal bar chart displaying only the ‘yes’ percentage only (Figure 5(B)).

(A) Horizontal Stacked Bar

(B) Horizontal Bar chart

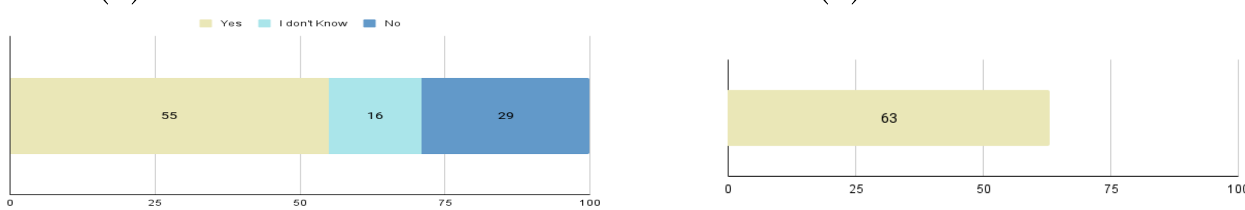


Figure 5: (A) Visualizing the student experience based on ‘Yeses’, ‘I don’t Know’, ‘No’ (B) Visualizing ‘Yeses’ of the student experience

Discussion and Conclusion

The main objective of this study was to provide insights from the design process of the Student Electronic Exit Ticket (SEET) system and teachers' thought process. We presented key teacher and instructional coaches’ experiences that played a pivotal role in finalizing the visual feedback displays. A rich understanding of student experience should support teachers to modify their instructional or curricular agenda to better promote equity in the classroom. Along with the help of teacher partners, our team has designed and developed an exemplar application supporting ‘Equity Analytics’. Shah and Reinholtz (2018) called for researchers to build tools and extend ‘Equity Analytics’ work with its applications to support equity across different disciplines. With the wide adoption of such tools, teachers can understand student experiences at scale to make classrooms equitable.

Our focus in this paper has been on analyzing how teachers perceive equity in the classroom when aided with visual learning analytics. We learned that it is imperative for researchers and system designers to have an open perspective when designing feedback dashboards and to deeply attend to teachers' perspectives and guidance in order to ultimately bring equity to the classroom in meaningful and relevant ways. The teachers' part of the study were from same school district and subject area (middle school science), a broader and more diverse group of teachers would be needed to further validate our findings.

Understanding student experiences can help teachers to shape classrooms to be more equitable by making them culturally relevant and sustainable (Paris & Alim, 2017; Brown, 2019), enabling all students to contribute and build on each other's knowledge. It is critical to quantify differences in students' classroom experiences in order to remove inequities based on characteristics such as gender and ethnicity. Ahn and colleagues (2019) discussed the value of using human-centered-design methods with teachers when designing instructor dashboards providing data on student learning. The aim of our design process was to apply such methods to explore how learning analytics tools can be adapted rather than adopted by teachers and to understand how the design process itself can foster new innovations in designing dashboards. Adapting these design methods influenced the design of the tool and helped with improvement efforts of supporting mathematics teachers. The dashboard helped to visualize patterns easily in data on experience for groups of individuals, such as individuals who are in the same classrooms in the context of their teacher-researcher partnership. Co-designing visual analytics with teachers can play an important role, providing help with interpretable data that enables them to make informed judgments about students' experiences and to modify their instructional practices accordingly. At the same time, preparing teachers to make sense of these data is not sufficient to promote equity, without consideration for systemic factors such as segregation, racism, and gender inequality in society--that help explain patterns of inequity in student experience data.

Chapter III: The Science Student Electronic Exit Ticket (SEET) System: A longitudinal study examining how different visualization forms support teachers to notice and reflect on inequalities in student classroom experiences

Ali Raza, Yamileth Salinas Del Val, William R. Penuel, Tamara Sumner, Melissa Campanella
University of Colorado Boulder

Abstract

This study examined the ways in which an equity analytics tool - the Science SEET system - supported middle school science teachers to reflect on the experiences of diverse students in their classrooms. The tool provides teachers with 'equity visualizations' - disaggregating classroom data by gender and race - that are designed to support teachers to notice and reflect on inequitable patterns in student participation in classroom knowledge-building activities, as well as 'whole class visualizations' that enable teachers to look at whole class participation patterns. The visualizations were based on survey data collected from students reflecting on day's lessons, responding to questions aligned with three theoretical constructs indicative of equitable participation in science classrooms: coherence, relevance, and contribution. The study involved 42 teachers, divided into two cohorts, participating in a two-month professional learning series. Diary studies and semi-structured interviews were used to probe teachers' perceptions of the visualizations' usability, usefulness, and utility for supporting their reflections on student experiences and instructional practices. A key result is that only the 'equity visualizations' prompted teacher reflections on diverse student experiences. However, despite the support equity visualizations provided for this core task, the teachers consistently ranked the whole class visualizations as more usable and useful.

Introduction

Students in public schools in the U.S. have become increasingly diverse across many dimensions, including gender, race, culture, and socioeconomic status (Digital Promise, 2011). In response to these changing demographics, teachers and schools are looking for approaches and tools to help them implement culturally responsive pedagogies and create equitable classrooms, where all students are able to participate in rich and engaging learning experiences. In our work, we define equitable classrooms as those where students' participation in and contributions to classroom knowledge-building activities cannot be predicted by their race, gender or culture. However, research has repeatedly shown that creating equitable classrooms can be very challenging for teachers. For instance, in science classrooms, multiple studies have documented the ways in which students experience inequity in their classrooms based on their race, gender, or the language they speak at home (Carlone, 2004; Warren & Rosebery, 2011; Wright, 2018). In small group work, differences in student status based on their race and gender have been shown to undermine their participation in collaborative knowledge building (Theobald, 2017).

Recently, there have been significant advances in the nascent field of "equity analytics", which explores how learning analytic paradigms can be used to help teachers create and sustain equitable classroom learning environments (Reinholz & Shah, 2018; 2021). Equity analytics applications systematically gather data about student participation in classroom activities and provide visualizations of these data, disaggregated by gender, race, and other important social markers. The rationale behind these systems is that teachers can reflect in these visualizations, notice patterns of student inequities in participation, and use this information to modify their instructional practices.

Here, we present a novel web-based application – the Science SEET System - which is specifically designed to support middle and high school science teachers to create more equitable classroom experiences. Teachers use this system to collect student experience data, in the form of short surveys, asking students how they participated or contributed to knowledge-building activities in class that day. The application provides teachers with several ways of visualizing their student data, including three 'equity visualizations' that have been carefully designed to support teachers to notice differences in student participation across race and gender, and three 'whole class visualizations' that depict student experience data for the class as a whole. The three equity visualization types emerged from a year-long, iterative co-design process with middle school science teachers that has been described elsewhere (Raza et al., 2020; Raza et al., 2021).

In this study, we examined how teachers used this novel equity analytics system 'in the wild', focusing on understanding how, when and why they choose to use, or not use, the provided equity visualizations. The two research questions motivating this study are: 1) When presented with visualizations that break down experience by race and gender, do teachers notice differences in race and gender? 2) Do teachers prefer visualizations that break down data by race or gender to whole class ones?

To answer these questions, we conducted a mixed-methods, longitudinal study with 42 middle school science teachers who used the Science SEET system over a two-month period. Our methods included a diary study, where teachers filled out a reflective questionnaire every time they used the system, as well as semi-structured interviews which probed them about their visualization preferences and their rationale for choosing different visualizations. Our study found that only the equity visualizations prompted teachers to notice and reflect on inequities in student participation based on race and gender. However, our study revealed that teachers greatly preferred the 'whole class visualizations' provided through the system to the 'equity visualizations', due to a variety of factors, ranging from their perceived usability of each visualization type, the teachers' own perspectives on equity, and the demographics of their particular classroom. As such, this study contributes to the field's knowledge and understanding on the utility of learning analytic dashboards for supporting diversity, equity, and inclusion in teaching and learning (Williamson & Kizilcec, 2021; 2022).

Theory & Related Work

This section describes the theories influencing the design of Science SEET system as well as related prior work on equity analytics systems and teachers' use of learning analytics dashboards and visualizations.

Constructs and instruments for characterizing student experience

At the core of any approach to developing an equity analytics system is deciding: (1) what constructs will be used to measure and characterize student participation in, or perceptions of, the learning environment and (2) how information on these constructs will be gathered. Reinholz & Shah, (2018) developed a classroom observation tool ("EQUIP") that relied on measures of student discourse, such as type of discourse (content-focused or logistics-focused), length, and wait time. In this work, a trained observer was sent to classrooms to score student discourse using the 'Equip' framework. These data were used in visualizations that enabled teachers to examine student talk during a particular lesson and over time, disaggregated by gender and race. Commercial applications are also emerging in this space. For instance, the PERTS system builds on constructs such as students' feelings of belongingness, the perceived relevance of their schoolwork, whether they received feedback from their teacher and other constructs (PERTS, 2010). In this approach, surveys ask students to reflect on the past week and answer a series of questions tied to these constructs.

In this study, we used three constructs to measure and characterize a student's classroom experience: coherence, relevance and contribution. Prior research indicates that these three constructs are a good predictor of classroom equity in science classrooms (Penuel et al., 2018; Penual & Watkins, 2019). Coherence refers to the students' perspectives of the progressions of their learning experiences, and whether these progressions were driven by student questions, ideas and investigations (Reiser, Novak, & McGill, 2017; Reiser et al., 2021). Coherence is particularly salient in contemporary science classrooms as it is a marker of the degree to which students are engaged in 'figuring out' a scientific phenomenon or developing solutions to a design problem. As noted by Reiser et al. (2021) "coherence arises when students see their science work as making progress on questions and problems their classroom community has committed to address, rather than simply following directions from textbooks or teachers." The relevance construct is rooted in expectancy value theory (Eccles, 1983), which highlights the importance of enabling students to make connections between their interests and identities and classroom activities. These relevance connections can lead to motivation, engagement, persistence, and self-regulation in learning (NASEM, 2018). We chose 'contribution' as our third construct as students' contributions in science classrooms as partners in knowledge building are considered critical in contemporary science reform (Schwarz., Passmore, & Reiser, 2017), as classrooms shift from a model where the teacher is the sole instructional, knowledgeable, and authoring agent. Rather, in science classrooms, the goal is to position students to be epistemic agents, i.e., "individuals or groups who take, or are granted, responsibility for shaping the knowledge and practice of a community" (Stroupe, 2014, p. 488).

We draw on experience sampling methodologies (Larson & Csikszentmihalyi, 2014) and use short surveys to ask students to reflect on the lesson they just experienced that day. Note that these surveys are not administered daily, but at the discretion of the teacher who selects specific lessons to gather feedback on. Each student responds to a short survey containing 10 questions delving into aspects of their classroom experiences that are aligned with each of the three constructs. For each question, students are asked to choose either “yes”, “no”, or “not sure”. Coherence questions include: 1) We work together to determine what ideas are most persuasive. 2) The teacher guides us to share our prior experiences or ideas about a phenomenon or topic to inform what we will do next. 3) Today we started class by reminding ourselves what we learned in the last class. Relevance questions include: 1) Today's science lesson was personally meaningful. 2) I found today's lesson interesting. 3) If people in my city or town understood the science we learned in today's lesson, they would do something that could help make our city or town a better place. Contribution questions include: 1) The teacher encourages us to build on and critique one another’s ideas. 2) Everyone’s ideas are heard. 3) Did you share any ideas out loud today to the whole class, a small group or a partner? 4) If you answered yes to the last question, did any of your ideas influence the class or help others?

Teacher Use of Learning Analytics Dashboards and Visualizations

Visualizations are a key component of many learning analytics dashboards designed for teachers. There has been extensive research examining how a variety of visualization types can support teachers to reflect on whole classroom activities; these visualization types include bar charts, line graphs, dot plot charts, lollipop charts, timelines, node-link diagrams, and heatmaps etc., (Ahn et al., 2019; Li & Wise, 2021; Dourado et al., 2021; Vieira, Parsons, & Byrd, 2018; Wise & Jung, 2019). Studies of these systems highlight the challenges teachers face when trying to interpret student data. Factors such as ease-of-use and perceived usefulness have been shown to impact dashboard use (Ali et al., 2013; Dazo et al., 2017). Recent research suggests that visualizations requiring a “learning curve” can impede teacher adoption of learning analytics dashboards (Arthars & Liu, 2020). These usability challenges can limit the kinds of inferences teachers are able to draw from their student data (Li & Wise, 2021); research suggests that as teachers become more proficient in dashboard use, they are able to draw more connections between their pedagogical knowledge and their classroom data (Molenaar & Knoop-van Campen, 2018). Another factor hindering teachers' effective use of learning analytics dashboards is the lack of alignment between what the dashboard is visualizing and the teachers' "pedagogical intent" or goals for their classroom (Lockyer, Heathcote, & Dawson, 2013). Despite this extensive body of research into visualization types for supporting teacher reflection, most visualizations don't support disaggregation of classroom data by gender or race for equity purposes (Williamson & Kizilcec, 2021; 2022; Vieira, Parsons, & Byrd, 2018). As a result, there have been efforts and calls within the learning analytics field to design tools for diversity, equity and inclusion to support learning for all learners (Wise, Sarmiento, & Boothe 2021; Ochoa, Knight, & Wise, 2020).

Understanding teachers' sensemaking processes as they use different dashboards and visualizations is a rich and critical area of inquiry within learning analytics (Verbert et al., 2020). Teachers' sensemaking processes often have three distinct stages: teachers' come to the visualization with an area of curiosity, they interpret data by 'reading' the visualization, and they generate explanations of the patterns they observe (Wise & Jung, 2019). While teachers may have an established area of curiosity, they do not necessarily come with predefined questions that they are trying to answer (Molenaar & Knoop-van Campen, 2018; Wise & Jung, 2019). Furthermore, different teachers will draw different meanings from the same visualization, which Ahn et al. (2019) refer to as the "one chart many meanings" phenomenon. In their study, this variability arose from differences in teachers' instructional practices and their classroom contexts. This body of work suggests that there are complex interactions between the types of data being visualized and the types of representations being used in the sensemaking process, necessitating further studies when introducing new visualizations, data types and representations. In this study, we extend this body of literature by examining how novel equity visualizations support teachers to reflect on their student experience data and how these data relate to their instructional practices.

Study Design and Methods

This study was conducted during a virtual professional learning series for middle school science teachers. The purpose of the professional learning series was to support teachers to improve their instruction by creating more equitable participation and learning opportunities in their classrooms. Teachers collaborated in small groups of 3 - 5, meeting four times over a two-month period. Each meeting lasted approximately two hours. All participating teachers used the visualizations and Science SEET system workflow described below throughout the series. Teachers received training on SEET system and visualization interpretation as part of the series. The professional learning series was conducted twice, consecutively, with two different cohorts of teachers.

Learning analytics visualizations & equity analytics tool

The Visualizations

The six visualizations used in this study render student classroom experience data organized around three constructs described in the previous section: coherence, relevance, and contribution. These visualizations were co-designed with secondary science teachers during a longitudinal, iterative design process described elsewhere (Raza et al., 2020; Raza et al., 2021). Three of the visualizations, which we refer to as the 'Whole Class Set', show only whole classroom responses for each of the three constructs. The remaining three visualizations we refer to as the "Equity Set", as they show disaggregated views of student classroom experience data broken down by gender and race.

Whole Class Visualization Set. Table 2 provides examples of each of the three whole class visualizations. The Overall Bar Chart (Table 2a) shows the percentage of students who responded ‘yes’ to one question about student experience as well as the number of students who attempted the question. It is important to note that this visualization is not a single bar chart, but rather 10 small bar charts displaying the response rate for each question in the survey. This approach of showing ‘small multiple bar charts’, builds on the work of Heer, Bostock & Ogievetsky, 2010. The Over time by Constructs (Table 2b) line chart displays the average of all questions corresponding to a single student experience construct. That is, this visualization shows three line charts - one per construct - with interactive legends that enables the viewer to selectively show or hide student responses associated with particular constructs. The Over time by Questions (Table 2c) line chart depicts change over time, by question responses, across multiple lessons. The y-axis shows the percent of ‘yes’ responses, and the x-axis shows the date. Questions belonging to a specific construct can be opened and closed by clicking on the legend. This chart also provides an interactive legend similar to the one described for the previous visualization.

Equity Visualization Set. Table 2 provides examples of each of the three equity visualizations, which are extensions of the visualizations described above with extra functionality to disaggregate student responses by race and gender. The Disaggregated Bar Chart by gender and race (Table 2d) shows data for one lesson. The y-axis shows the demographic data and the number of students, while the x-axis shows the percentage of students responding ‘yes’ to the particular question. As described above, this visualization shows 10 of these bar charts corresponding to each question. The Over time by Gender and Race (Table 2e) line chart shows the percentage of 'yes' responses to all of the student experience questions for a specific demographic marker such as female, non-binary, or African American. The Choropleth Heatmap (Table 2f) displays gender and race data on the y-axis. A color encoding is used in each matrix to represent student data: when more students respond 'yes' to a question, the color intensity will be higher. When hovering over any cell in the matrix, a tooltip will display the question, percent 'yes' answer, and the number of students responding.

Teacher and Student Interactions with the Science SEET system.

Teachers use the SEET system to collect information about students' perceptions of their classroom experiences during a specific lesson. The high-level workflow is shown in Figure 6. The workflow is initiated when teachers administer a student experience survey (Step 1) to a particular classroom and students respond (Step 2). The 10 questions used in these surveys and their underlying constructs were described in Section 2. Student responses are automatically tabulated and visualized to support teacher sensemaking (Step 3). In the fourth stage, teachers select new instructional strategies to implement in their classrooms (Step 4), and the cycle begins anew. Before using the system for the first time, students are explained the purpose of SEET system and how their data will be used, by their teacher and with a recorded video prepared by the research team.

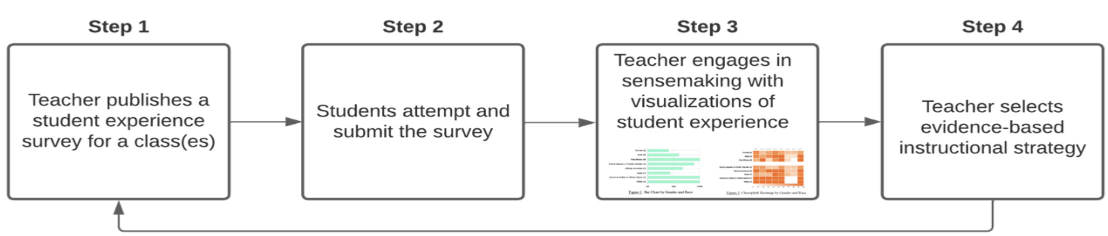


Figure 6: Teacher workflow using the Science SEET system

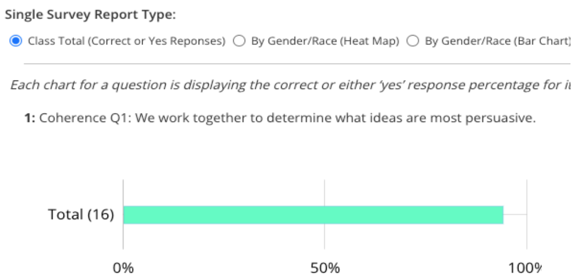
Student Anonymity

This equity analytics system is carefully designed to protect students' anonymity so that they feel comfortable providing their teachers with this feedback. The system generates a unique, one-off identification code for a student to use that school year. Students enter their demographic data, such as race, gender, and language at home, once when they sign up for that class. Thus, there is no tracking of students across courses, over time. To further protect anonymity, the system does not report results for groups of 2 or less (i.e., female or African American, etc.).

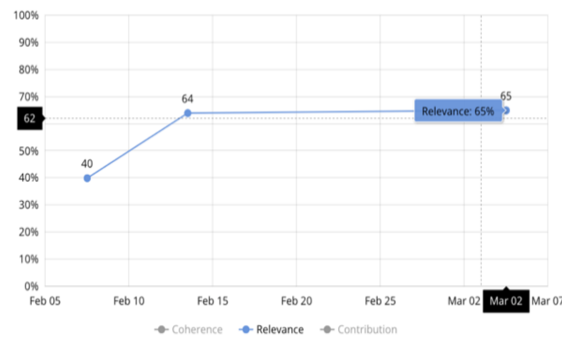
Participants

Two cohorts of middle school science teachers were recruited for spring and fall semesters in 2021. There were no overlapping teachers in these two cohorts. Teachers across the United States were recruited using social media (Twitter) and through professional organizations such as the Council of State Science Supervisors. For this study, we are treating these cohorts as one population as we are focused on their use and interpretation of the visualizations, which were the same across the two professional learning series. This study population consisted of 42 teachers (33 female, 9 male) from 19 different U.S. states. The teachers were majority white (27), with the remaining 15 teachers identifying as African American or Black (2), Asian American (4), Latinx or Hispanic (2), Native American (1), or Other (6). Three of the teachers had less than 2 years experience; the average teaching experience was 12 years. All teachers were compensated at an hourly rate for their participation in the workshop series and research data collection, averaging between $300 to $500 per teacher.

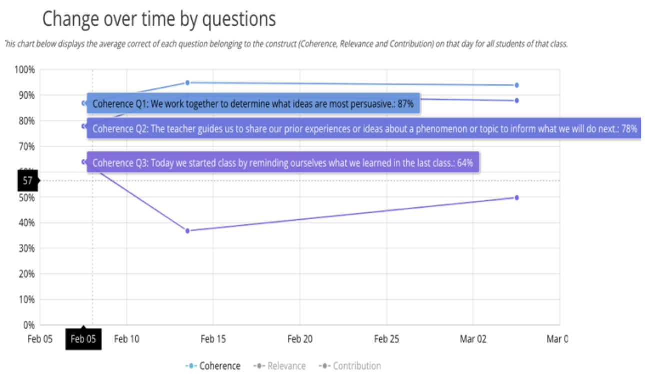
Table 2: Whole class visualizations displayed in 2a, 2b, and 2c. Equity set visualizations displayed in 2d, 2e, and 2f.



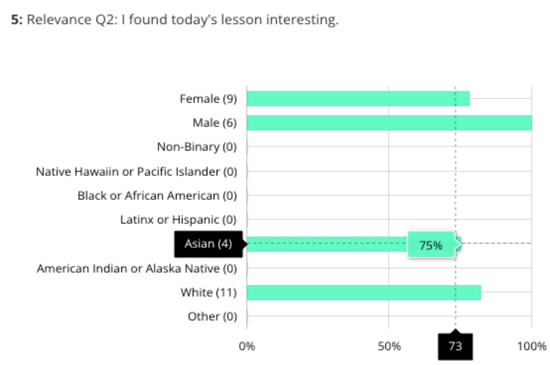
2a: Overall Bar Chart



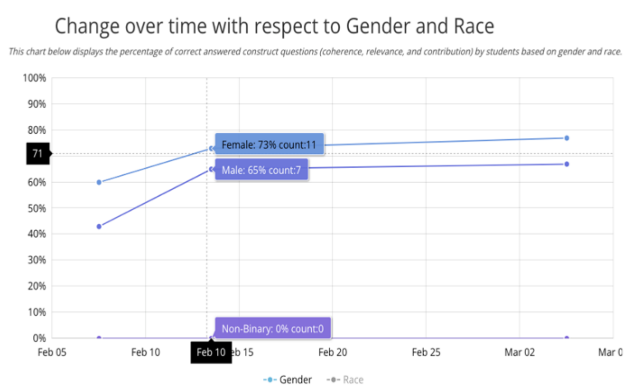
2b: Over Time by Constructs



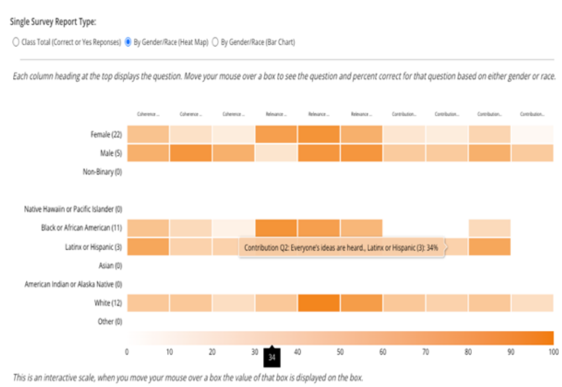
2c: Over Time by Question



2d: Disaggregated Bar Charts



2e: Over Time by Gender and Race



2f: Choropleth Heatmap

Data Collection

The two data sources used in this study were diary study entries and semi-structured interviews.

Diary Study. The diary study method asks participants to systematically record qualitative or quantitative data about their experiences, building a "diary" as they log individual entries over time (Diary Study, 2016). In this study, teachers were asked to use SEET system and complete diary entries: (1) the first time they collected and viewed their student experience data, and (2) after each time they tried a new instructional strategy in their classroom. Teachers made entries into their diaries throughout the two-month professional learning series. Each teacher's 'diary' consisted of an anonymized folder containing a shared Google document, where this document provided a template to be used for each diary entry. The template was divided into areas that posed different questions to prompt teacher reflections, such as "What patterns do you notice in the data, and what do you wonder about?" To answer this prompt, we provided a three-column table with the following headings: Things I notice, Things I am wondering about, Which visual representation I am looking at for patterns. For full template, see Appendix B.

Semi-structured Interviews. We conducted a semi-structured interview ranging from 10-15 mins with each teacher after the conclusion of the professional learning series. These interviews were conducted online using the Zoom video conferencing tool, and all interviews were video-recorded and transcribed. During the interview, the researchers shared their screen and showed the interview participant a figure containing all six visualizations. Participants were asked to "rank each visualization in relation to how useful they were to your practice, from most to least useful, and also provide a rationale for your rank". After each participant responded to this prompt, the researcher repeated the rank the teacher had assigned to a particular visualization and allowed the teacher to change their ranking if so desired.

Data Analysis

We analyzed data collected from the diary study to answer our first research question: do teachers notice differences in race and gender when examining equity visualizations. We merged all of the teacher reflections from their individual diary entries into one data source. We then organized these reflections by visualization type; i.e., the 6 visualizations described earlier. One of the authors then analyzed these reflections to see which visualization prompted 'noticings or wonderings' related to equity and/or one of the three constructs (coherence, relevance, contribution). To count as an equity-related reflection, the teacher needed to explicitly refer to students' race or gender in their diary entry. Similarly, to count as a construct-related reflection, the teacher had to explicitly mention one of the three constructs in their diary entry.

To answer the second research question - do teachers prefer equity visualizations over whole class ones - we turned to the semi-structured interview data. Because teachers were being asked to reflect on each visualization, we cleaned and merged the interview data from each teacher into a visualization-specific data source, one for each visualization type. To analyze these data, we used inductive coding following a grounded theory approach (Charmaz, 2006). Two authors conducted an initial analysis and created six code books, one for each visualization type. One of the authors wrote code definitions and shared them with the other coder. Both authors coded transcripts line-by-line. The two authors then met multiple times to reach reliability in any category with a percentage agreement of less than 80 percent; disagreements were discussed until consensus was reached (Miles & Huberman, 1994). After finalizing the inductive code books, the two authors met multiple times to refine and review the categories and to create cross-visualization themes. This approach enabled us to identify both visualization-specific attributes that influenced teachers' rankings as well as general themes (not specific to any particular type) that influence teachers' rankings. To ensure reliability and validity, the team followed recommended best practices for qualitative data analysis (Miles & Huberman, 1994), including keeping an audit trail of all codes, reaching high agreement between two coders to reduce bias, documenting all decisions made during the analytic process, and involving authors other than the coders to inspect alignment between data and interpretations.

Results

Our results are organized around the research questions posed in the Introduction.

RQ1: Do teachers notice differences in students' classroom experiences when presented with equity visualizations?

In this study, teachers recorded 248 "noticing and wondering" reflections in their diaries. Of these 248 reflections, 35 considered students' race and/or gender, while 213 reflections considered the three constructs of student experience (coherence, relevance, and contribution). As shown in Table 3, only the equity visualizations prompted teachers to reflect on how students' race or gender may impact their classroom experience. Teachers reported zero reflections related to race and gender when using the whole class visualizations. Both sets of visualizations - equity and whole class - prompted reflections on student experience related to the three constructs, with the whole class visualizations prompting 76% of the construct-related reflections compared to only 24% by the equity visualizations.

Table 3: Teacher noticing and wondering with different visualization types

Diary Study Reflection	Equity set	Whole Class	Total reflections
Related to race and gender	35 (100%)	0 (0%)	35
Related to constructs	52 (24%)	161 (76%)	213
Interview Reflection			
Related to race and gender	71 (92%)	6 (8%)	77

These excerpts from teachers' recorded reflections illustrate how these two types of reflections differ based on visualization type. For instance, when Teacher13 was interacting with the Bar Chart Disaggregated by Race and Gender, they noticed: "*I am not surprised that the girls ranked lower in all three of these questions (contribution). Although there are a significantly higher number of girls in this class, the boys tend to speak up more, share more, discuss more, etc*." This contrasts with the reflections of Teacher10 when interacting with the whole class Overall Bar Chart: "*Students are still not able to make connections of the concepts they are learning with real life. Students still feel that their ideas do not influence the class or help others*." In this reflection, the bar chart prompted the teacher to think about relevance and contribution for the students as a whole.

We saw similar results in the interview data, where equity set visualizations prompted the majority of reflections on race and gender (92%), compared to the whole class visualizations (8%) (Table 3). All three of the equity visualizations prompted reflections on gender and race during the interviews: Over Time by Gender and Race (54%), Heat map (21%), and Disaggregated Bar Chart (17%). Interestingly, the majority of equity reflections focused on gender (29 out of 42 teachers, or 70%), while only 16 out of 42 (38%) of teachers commented on race. Teacher29 talks about this disposition to focus on gender in her interview "*Oh, I think it does a lot for me actually … I'll focus more I guess the gender than the race, since I do have more of one student, but the gender it really helps me understand Okay, am I, like tailoring my teaching more towards the boys or am I, you know, do* *the girls feel like they're not being heard. Right like maybe, am I hearing more boys' voices and my calling on them more than the girls, etc, so I think that I don't ever really pay attention to that um and so just having the data in front of me I can be like Okay, let me be intentional about what I'm doing now*." The interview data also highlighted some of the ways in which the equity visualizations supported teachers to reflect on patterns in their classroom, with teachers reporting that some visualizations helped them to view how specific groups participated in class (by gender and race) or supported them to view intersectionality across groups (for instance, Black girls). Teacher40 talks about the value of the Heatmap for supporting their equity reflections: "*As time went on, I began looking more at the heat map and I was using the heatmap to see which students were responding in which way. So is it mostly you know how do Hispanics see themselves how do white see themselves how do the females and males and the Asians, how did they all see themselves to perceive themselves within the classroom and what is it that I can do to quickly tweak my teaching in order to see changes in the Heatmap*."

Some teachers also mentioned that looking at equity visualizations featuring gender and race could be 'emotionally hard' to make sense of. For instance, Teacher 6 described that he did not like the Disaggregated Bar Chart because it forced him to look at intersectional data combining gender and race "*…. I think, in my brain, I find it harder to kind of combine the idea that, for example, like it's disaggregated in like you know he's seeing that. You see, like the total for gender and then you see the total for race and like that can be kind of a little bit of dissonance in my brain. … like, for example. or Black girls I'm not doing you know enough so I think maybe that's why I gravitate towards that one least*."

RQ2: Do teachers prefer visualizations that break down student experience data by race or gender to the whole class ones?

To answer this question, we analyzed data collected from interviews with teachers, asking them about their use of the different visualizations available in the system. As part of the interview, teachers were asked to rank order the six visualizations based on the amount that they used them, where a ranking of 1 corresponds to their most used visualization and a ranking of 6 is least used; they were also asked to provide rationale justifying these rankings.

As shown in Table 4, teachers ranked all the whole class visualizations higher than any of the equity visualizations. This preference for the three whole class visualizations held across teachers' individual rankings of visualizations as well as the sum of rankings across all the teachers.

Table 4: Teachers' rankings of the six visualizations

	Whole Class visualizations			Equity Set visualizations		
	Over Time by Constructs	Overall Bar Chart	Over Time by Question	Disaggregated Bar Chart	Over Time by Gender and Race	Choropleth Heatmap
Median rank	2.5	3	3	4	4	5
Sum of rank	102	98	110	125	130	148

To better understand why teachers ranked visualizations in a particular way, we coded the rationale they provided in their interview for each ranked visualization. Our codes examined teachers' perspectives on the ease-of-use of particular visualizations, whether teachers indicated that the visualization was useful for reflecting on equity, and their perceptions of the utility of the visualization for their classroom context. Individual codes were then grouped into higher-level themes (Table 5).

The Choropleth Heatmap was ranked the lowest by teachers, while relatively simple visualizations such as the line and bar charts were ranked the highest. This preference for simple encodings is underscored throughout their interview data. For example, Teacher 1 ranked the 'overall bar chart as her first preference, noting "Bar charts you got to physically see the zero percent 50% hundred percent you could watch the numbers change from question to question… the bar chart, in my opinion, is the easiest to read and the fastest to get the information you want from it and it's the simplest." Our qualitative coding of the interviews found that many teachers expressed similar opinions, citing the bar chart's overall ease of use: simple/easier (24), helpful/quickest (9), faster to get information (6), feels confident in reading (4), likeness (2), Good insight (3).

Teachers reported that both of the whole class "Over Time by" line charts were easy-to-use, with the Over time by Construct visualization supporting them to easily spot high level trends, while the Over time by Question visualization better supported exploration of the data. For instance, Teacher3 describes their use of the Over Time by Construct visualization (ranked #1) to spot trends: "*I like looking at the lines. For me, it was just a quick visual of the directionality of where I was heading with my data, And then I could really dig in more*." Whereas Teacher14 describes how they used the Over Time by Questions visualization (ranked #1) to explore their classroom data: "*….the one that was most useful to me or the one that I spent the most time looking at was overtime by question. Because that allowed me to see growth over the three data collection points or times and how specific questions were associated with that, as opposed to overtime by construct, which is where you had questions that are grouped together that didn't tell you necessarily exactly what the responses were so I chose overtime by question as the first one*."

When reflecting on her use of the heatmap (ranked last, #6), Teacher2 remarked "*I could tell you I don't like the heatmap. All the rest are great. I love others…. but I don't like the heat map. I must not be very smart, [….]. Maybe, because my data is not as big. I just don't feel like I didn't get anything out there,* *that I couldn't get out of the other charts, but that's just me. Because I don't know how to use it.*" Frustrations with the Heatmap visualization were evident across many teachers, with others citing issues such as it is hard to read or make sense (17), hard to interpret the colors, difficult to remember the meaning of the colors, too many colors (13), no previous experience/familiarity (17), and a high learning curve (8).

It is important to note that not all the teachers found the Heatmap difficult to use. Teacher3 ranked it highly (top rank of 1). He justified this high ranking by explaining that it provided all the information he wanted in one place and he had prior experience with similar visualizations: "*Because it's like everything in one place and it's really easy to make really targeted decisions there. I wonder about like yeah like you know we all come from really different backgrounds and like pretty data as I used to be in geospatial science pretty comfortable with kind of visualizations like that…*". A few other teachers also found value with this type of visualization, indicating the Heatmap was useful for: helpful in displaying gender/race (8), overall/summarizing (6), liked color shading (5), useful tool (5), easy for targeted decisions (4).

Table 5: Themes influencing teachers' rankings of each visualization

Visualization	Theme name (Code frequency)
Overall Bar Chart	Visually easy to read (30); Supportive (9), provide overall classroom data (22), Help with starting analysis (15), Helpful in seeing one question at a time (14)
Over Time by Construct	Visually easy to read (29); Helps inform teaching (11), provided an overall view of student experience (15)
Over Time by Question	Visually easy to read (11); Too much going on (10); Facilitates granular look at data (36), did not find the need to use (4)
Disaggregated Bar Chart	Easier to read with confidence (22); Facilitates granular look at data, including equity-oriented data (18); Visualization not useful (11); Data not useful (3)
Over Time by Gender and Race	Mixed perceptions of readability (6); Facilitates granular look at data, including equity-oriented data (57); Helps inform teaching (9); Data not useful (15); Diverging perspectives on equity (4)

Visualization	Theme name (Code frequency)
Choropleth Heatmap	Complex to read and interpret (41); Little familiarity with this chart type (26); Facilitates granular look at data, including equity-oriented data (18); Helps inform teaching (6); Diverging perspectives on utility (21); Data not useful (4)

Perceived ease-of-use based on simplicity and familiarity

When providing rationale as to why they preferred visualizations such as the whole class bar and line charts, teachers cited reasons based on the simplicity of the representation and/or their prior familiarity with the representation. For example, when explaining why they liked both of the provided bar charts (Overall and Disaggregated), teachers cited reasons such as simple/easier (24), helpful/quickest (9), faster to get information (6), feels confident in reading (4), familiarity with bar chart helped with disaggregated bar chart (3), likeness (2), not too much data (2), and past usage (1). For instance, Teacher10 cites her familiarity with bar charts as the reasons she ranked the Overall Bar Chart 2nd: "*I usually use the bar charts in my classroom … But my classroom has one wall, the coordinate plane. You will see all the coordinate planes on one wall and the other one has exactly like you know their goals….. And then I show them like a bar. …, so I think, because I have been talking to them [her students] about bar charts a lot so maybe that's the reason*." Similarly, Teacher7 explains why she rated the Disaggregated Bar Chart 2nd, behind the Overall Bar Chart that she ranked first "*I would say the next most helpful would be the disaggregated bar chart. I'm just a bar chart person. That was the simplest*."

Teachers' perspectives on the ease-of-use of the Choropleth Heatmap were polarized. A few teachers (12) found the Choropleth Heatmap easy-to-use, often citing their prior familiarity with these types of representations. In these cases, teachers reported that the use of color to denote the percentage of students responding "yes" positively contributed to their practice by providing a quick snapshot of the day's lesson across equity markers (race and gender). In their interviews, they described the Heatmap in the following ways: supporting their overall/summarizing (6), liking the color shading (5), helping them to identify patterns (3), For instance, Teacher33 ranked the Heatmap 3rd, noting that it provides a quick overview of student experience: "*… I think usability it's really easy to see based on the color quickly, sure that was really interesting. I just thought it was fascinating to see either how students responded similarly or responded differently to the same question*." Most teachers, however, reported finding the Heatmap very difficult to use, often citing the use of color in this representation as being problematic (Figure 2f). For instance, teachers reported that the colors make the Heatmap representation hard to read or make sense of (17), colors make interpretation difficult, it is difficult to remember the color codings or there are too many colors (13), too much data in one display (2). For example, Teacher3 narrates why she ranked it 6th: "*I don't know, my mind just doesn't think in this way of the heat maps. yeah I couldn't tell you exactly what it is, but it's just, it's there's so much stimulating the mind, I think, is what it is*."

Barriers to using equity visualizations

When talking about why they did not prefer the equity visualizations, two themes emerged from the teachers' interviews: one related to their classroom context and one related to their own perspectives on equity. With respect to their classroom context, some teachers (13) stated that there was little reason for them to look at visualizations disaggregated by race and gender as there was very little racial or gender diversity in the students in their classrooms. Teachers noted that in some cases, there may have been only 1 or 2 Black students, or one transgender student, in their classroom. In such cases, the SEET system would not have displayed this information due to the protocol for preserving student anonymity, which prevents the system from displaying data when there were 2 or fewer students in a group.

We see evidence for this concern in the teacher interviews across all three equity visualizations: less useful due to lack of sample (4), not a large enough sample size (7), small sample size made it less useful (3), not all students were included in race (1). Consider the excerpt below when Teacher5 talks about how limited gender diversity restricted her use of the Disaggregated Bar Chart, ranked 6th "*I just think and for me, you know disaggregated I'm looking at. You know it's female and male. There weren't very many kids that identified as non-binary. Maybe only had one or two and then it's just the white and I just don't have like if I look here (pointing to other races)*". Similarly, Teacher24 talked about how lack of classroom diversity limited her use of Over Time by Gender and Race visualization, ranked 3rd. "*I think it's just really not applicable for mine or by race, I mean you know, so I just have such a homogenous group that the race wasn't there's really nothing to look at there…*"

The second barrier to use stemmed from teachers' own equity beliefs. Specifically, there was a small set of teachers (7) that reported that showing classroom participation data disaggregated by race and gender did not reflect their personal interests or their personal beliefs about classroom equity. For instance, Teacher19 explains that she was not personally interested in viewing one of the equity visualizations, noting "*Over time by gender…I just like that I'm not personally interested in that*." Teacher9 cites differences in personal beliefs, describing how she views equity through the lens of equal access to resources, "*...it's equality, we are talking about equity. Where does this thing affect gender and race? The public school system is providing equal opportunity. There are the same table, chairs, same teachers. Everything is the same. The teacher is not changing their methods. They are putting every individual, a need that you did not learn this question, let me try it this way, this way. So where is the concept of gender and race*." These perspectives contrast with those expressed by many teachers who noted the value of understanding student participation disaggregated by race and gender. Teacher15 discusses the value of looking at data reported by gender, "*But I will say gender plays more important role. Because in, especially in science and math classes. I have seen that male students, they try to dominate they try to ignore female voices. And females are usually not chosen by the group as the leaders*." While Teacher8 discusses how they value looking at classroom data disaggregated by race, "*I always want to see how my Latinos and Everybody in general, like, I feel that, like white people also benefit from seeing people of color excel you know, like it helps build their understanding that we are equals. You know what I mean and that's a good thing.*"

Discussion

In this study, we examined how middle school science teachers used visualizations, investigating teacher perceptions of the usability, use, and utility of these visualizations, as well as the factors that influenced teachers' perceptions. Our core finding is that the equity visualizations functioned as intended; that is, these three visualizations prompted teachers to reflect on how student participation patterns varied by race and gender. This finding contrasts with the whole class visualizations, which rarely prompted such equity-oriented reflections. Our second finding is that the majority of the teachers greatly preferred the whole class visualizations, for a variety of reasons, and many teachers reported using the whole class visualizations much more frequently. This is despite the fact that all of the participating teachers volunteered to participate in a longitudinal, two-month professional learning series in order to improve the equity of student participation in their science classrooms. Teachers reported that ease-of-use, simplicity, and familiarity with common chart types - such as bar and line charts - heavily influenced what visualizations they chose to use.

While this study shed light on our two research questions, it has also raised intriguing new questions and opportunities for the learning analytics field when designing equity analytic dashboards for use by K12 teachers.

Better support for teachers' data visualization literacy

We observed that teachers underutilized the equity visualizations, preferring the ease-of-use provided by the whole class visualizations. From a design perspective, we need further investigation into new visualization types that embody simplicity and familiarity, while still providing access to finer-grained data on students' participation by race and gender. However, equity visualizations necessarily embody more complex data types, suggesting that new visualization literacies may be needed. For instance, we used color in the Choropleth Heatmap to show student responses patterns. There was a polarity in teachers' responses to this use of color, with the majority of teachers disliking this use of color while others reporting it to be very helpful to their sensemaking. These differences appeared to stem from prior experience and training in different chart types. Thus, future research might focus on how to design learning opportunities for both pre-service and in-service teachers to prepare them to work with visual learning analytics tools. Prior research also suggests that past personal experiences with visualization types can influence the trust people have in them (Peck et al., 2019). Trust is likely an important attribute for visualizations displaying sensitive data by gender and race, raising a question for visualization designers: how does the usability of equity visualizations influence trust in teacher practitioners and other education stakeholders?

Specific supports for racial and gender sensemaking

There are extensive calls in the literature to focus on equitable learning experiences with visualization dashboards centered on diversity, justice, and equity (Williamson & Kizilcec, 2021; 2022; Wise, Sarmiento, & Boothe 2021). Teachers often cited a lack of racial diversity in their classrooms or their own perspectives on equity as reasons for not using the equity visualizations. Yet, in all cases, the students in these teachers' classrooms included a mix of genders, i.e., there were plenty of opportunities for teachers to use these representations to pursue gender equity goals. One line of research is looking at how to embed equity visualization dashboards into teacher preparatory programs (Shah & Coles, 2020) in order to build teacher capacity to engage in these types of sensemaking. Another potentially fruitful line of research could study the types of 'routines' that teachers could systematically enact to support their sensemaking. Routines - a repeatable sequence of instructional moves - are a common feature in contemporary science curricula (OpenSciEd, 2019) that support teachers to reliably enact knowledge-building activities in their classrooms. We can imagine the creation of supporting routines going hand-in-hand with the design and deployment of new equity analytic systems.

Limitation

This study only investigated the use of a narrow set of visualization types (Bar charts, Line charts, and Choropleth Heatmaps) organized around a specific type of student experience data (coherence, relevance, and contribution) disaggregated by race and gender. Future work can explore how other equity-oriented visualization types with novel student experience data can impact teachers' sensemaking.

A second limitation of this study relates to the lack of racial diversity in many of these teachers' classrooms. Many teachers cited that they did not use the equity visualizations because there were no patterns in student responses across races to examine. In this study, classroom diversity was not required for teacher participation. In future studies, we are instead partnering with school districts with diverse student demographics to ensure that participating teachers have racially diverse classrooms.

Conclusion

In this paper, we investigated teachers' usability, use, and utility of equity focused visualizations of student classroom experience data compared to whole class ones. We found only equity visualizations disaggregating data by race and gender prompted teachers' perceptions of their classroom equity. We also found several barriers to teacher use of these equity visualizations, such as ease-of-use, familiarity, the lack of diversity in their classrooms, and teachers' personal perspectives on equity. This longitudinal study contributes to the nascent field of designing for social justice in learning analytics, focusing on the ways in which an equity visualization dashboard can aid teachers in their workplace practice. In our future work, we aim to further develop and study the accompanying professional learning series, and to examine how these professional learning series and the Science SEET system can be embedded within school district infrastructures. We also intend to study how instructional support leaders within school systems - such as coaches or district leaders - can support teachers in using these equity visualizations to inform and guide new instructional practices.

Chapter IV: Supporting Science Teachers in Using Student Experience Data to Support More Equitable Participation in Science Classrooms

Ali Raza[1], Jiawei Li[2*], William R. Penuel[1*], Kenneth Frank[2]
[1]University of Colorado Boulder, Michigan State University
*Co-equal second authors

Abstract

Recent science education reforms increasingly aim to center equitable instruction that support learning experiences of all students in the classrooms. To this end, we investigated whether engaging teachers in a cycle of collaborative inquiry related to strategies for promoting more equitable participation in science could improve the quality of student experience for racially minoritized students and for students of different genders. We partnered with forty-nine middle school science teachers remotely in two cohorts over a course of one year in the Student Experience Inquiry Cycle (SEIC) and gathered data from student experience surveys to assess changes to student experiences. We analyzed the data from an explicit equity stance: seeking to promote equality of experience and, where possible, stronger gains for students most excluded from STEM. Findings show that SEIC supported improvements in the quality of students' experience for both teacher cohorts and in three cases, differentially benefited Native Hawaiian/Pacific Islander, Asian, and Black students. We conclude that formative use of student experience data can improve the quality of student experience in classrooms but that achieving equity goals is more challenging.

Supporting Science Teachers in Using Student Experience Data to Support More Equitable Participation in Science Classrooms

Research in science education and other disciplines has long underscored the value of formative assessment for learning. When informed by adequate theories of how students learn and supported by strategies that help link students' everyday forms of thinking to disciplinary forms of thinking and doing, formative assessment is a powerful tool for advancing disciplinary learning goals (Penuel & Shepard, 2016). Evidence from programs of in-service professional development, moreover, shows that teachers can grow in their formative assessment practice in ways that change classroom practice and improve student learning (Furtak et al., 2016; Penuel et al., 2017).

One largely neglected aspect of research on how formative assessment can support learning is the assessment of how students experience their learning environments. This gap is significant because research on learning emphasizes the need to involve students as partners with each other and the teacher in knowledge-building activities in the classroom (Engle, 2012; Stroupe, 2014; Tabak & Baumgartner, 2004), processes that are rarely directly measured in teachers' classroom assessments. Further, several studies in science education and other disciplines point to the ways that students' opportunities for knowledge building in the same classroom are experienced differently, depending on their race, their gender, and the languages they speak (Krumm et al., 2020; Langer-Osuna, 2015; Munter & Haines, 2019).

Our research has shown that it is practical for teachers to collect and interpret data on student experience to identify inequities in student experience linked to gender and race (Raza et al., 2020), but we have limited evidence that teachers can use such data to improve the quality of student experience in science classrooms overall and in a way that makes experience more equitable. In the current study, we the authors of this paper in partnership with other researchers and teachers worked with two cohorts of middle school science teachers in a design study to address the questions of how student experience varies and whether a professional learning cycle could support teachers to use student experience data formatively collaboratively as part of the cycle of inquiry to increase equity with respect to students' experiences of and contributions to knowledge-building activities in the classroom. To answer this question, we relied on evidence from student experience surveys collected before, during and after teachers adapted and enacted strategies to promote more equitable student experiences in their classrooms. We fit multilevel models to data to explore overall growth and to investigate whether students from racially minoritized groups and students who identified as female or gender non-binary benefited more from their teachers' participation in the cycle.

Our work was with teachers and our quantitative analyses were guided by specific equity commitments held by the team. The first is that, at a minimum, educators should strive toward equality in student experience, meaning that we should not be able to predict the race or gender of a student from knowing how they rated the quality of their experience of science learning. We also encouraged educators to adopt what Reinholz and Shah (2021) refer to as a logic of reparation: where possible, they should seek to center the experiences of racially minoritized students in deciding how to focus their own improvement efforts, recognizing that racism, sexism, and transphobia are systemic oppressions. In analyzing data, we have sought to decenter the perspectives of dominant groups (white- and male-identified students) and to call attention to the different sources of variation in student experience at the occasion, individual, and classroom level, so as to draw attention to both pattern and variation within groups in results (cf., Gutiérrez & Rogoff, 2003).

Conceptual Framework

A key focus of our effort was to help teachers gather evidence related to students' experience of classrooms. As we define it here, experience describes transactions between persons and settings within organized social activities, transactions that are suffused with affect (Roth & Jornet, 2014). Experience, in this sense, is not something an individual possesses (i.e., as "having an experience"), but rather is a transaction that unfolds in time as part of ongoing activity, with others, and within and across settings. An account of experience, moreover, requires some explanation of how events such as science lessons are produced within larger forms of organization (e.g., curricular structures, reform discourses) and how those give rise to particular social interactions (Roth & Jornet, 2014). Further, an account of experience needs to consider that the same setting is likely to be experienced differently by different people, depending upon the history of past transactions with others in it and how they feel about those interactions (Vygotsky, 1994).

Student Experience in Practice-Focused Science Classrooms

Inspired by images of science taken from the field of science and technology studies (e.g., Latour & Woolgar, 1986; Pickering, 1995), science education has undergone a sharp "practice turn." The practice turn reflects the view that learning goals should build students' proficiency in the practices of science, including its conceptual, social, epistemic, and material dimensions (Windschitl & Calabrese Barton, 2016). *A Framework for K-12 Science Education* (National Research Council, 2012), the Next Generation Science Standards developed from the framework (NGSS Lead States, 2013), as well as subsequent guidance to teachers regarding how to evaluate curriculum materials (Achieve, 2016) and assessments (Achieve, 2018) all reflect the goal of developing among students both an understanding of and grasp of how science practices can be used together to construct and critique answers to questions about the natural world.

Achieving the goals of a science-as-practice approach very much depends on students' experiences within science classrooms. In the vision of the *Framework*, students engage in posing and answering questions about phenomena (Berland et al., 2016), applying care in the context of solving socioecological problems in engineering design (Gunckel & Tolbert, 2018), and grappling with uncertainties with respect to how to design investigations to answer questions about the world (Manz, 2019). Importantly, these classroom science practices are often imagined as collective forms of activity that support a shared sense of collective enterprise or joint knowledge building (Alzen et al., 2020; Carlone et al., 2011; Manz, 2015), as well as giving students a "feeling for the discipline" (Jaber & Hammer, 2016) and preparing students to treat one another and their ideas with care (Krist & Suárez, 2018).

Of particular importance in practice-focused science classrooms are opportunities for students to contribute ideas toward building explanatory models of phenomena that students are investigating. Here, phenomena refer to "observable events that occur in the universe and that we can use our science knowledge to explain or predict" (Achieve, 2017, p. 1). As focal points for organizing instruction, phenomena allow teachers to present science not simply as a body of facts to be memorized, but rather as ideas and practices for making sense of the world around us (Berland et al., 2016; Schwarz et al., 2017). For their part, to gain a feeling for science as it is practiced, students need to make substantive contributions to explaining the phenomena at hand, in coordination with other students' ideas and with ideas the class has established so far as knowledge that can be built upon (Reiser et al., 2017).

Past research suggests that, even within classrooms where teachers are using curriculum materials and teaching in ways that support students' meaningful engagement in classrooms, there can be considerable variation in how students experience those classrooms. Some studies have documented wide variation across classrooms, for example, in terms of how much students experienced excitement in classrooms where students' teachers were using phenomenon-based curriculum materials (Carlone et al., 2014; Carlone et al., 2011; Morozov et al., 2014; Penuel et al., 2016). Research has also documented significant variation in student experience within classrooms within similar kinds of classrooms. For example, one study found that students' race and gender related to students' reports that they contributed ideas to discussions and that their ideas were heard (Krumm et al., 2020). Another found wide variation in how relevant students perceived the lesson to be personally; this study also found that whether they perceived the day's lesson was personally relevant to them was associated with the degree to which they felt a sense of ownership over their learning (Zivic et al., 2018).

In classrooms that we define as equitable, we would expect that we could not predict the race or a gender of a student by knowing whether they contributed ideas toward explaining phenomena and perceived that their ideas influence the course of investigations and explanatory modeling activities. Teachers in such classrooms would position students both as knowledge producers and users of science knowledge produced by peers and others, thus cultivating a sense of ownership over science ideas (Calabrese Barton & Osborne, 2001; O'Neill, 2010). Teachers would hold all students accountable to contributing their ideas to the class, while being given appropriate support and recognition from both peers and the teacher to do so (Carlone et al., 2011). Students, for their part, would each perceive themselves as enacting epistemic agency within the classroom, that is, as sharing ideas on the public plane of the classroom that are helping to build knowledge that is important to the classroom community (Stroupe, 2014).

Promoting equity may require going further, toward reparation of epistemic harms students experience in schools. Students who identify as female, nonbinary, or trans, as well as students who come from Black, Indigenous, and People of Color communities face what Fricker (2007) calls epistemic injustice society, that is, the regular discounting of their knowledge because of their (marginalized) social identity. Epistemic injustice shows up in classrooms and other sites of learning as well, when students' everyday ways of speaking, relating, and knowing are judged to be unwelcome in the classroom (Stroupe, 2021). Addressing it may require the adoption of an ideal of reparation (Reinholz & Shah, 2021), that is, of the need to center and address the experiences of those students most subject to harms of epistemic injustice.

Generating and Using Data on Students' Experience to Support More Equitable Experiences Using a Student Experience Improvement Cycle (SEIC)

While several scholars have proposed strategies for establishing more equitable classroom cultures and documented instances of such cultures in individual studies (Cohen & Lotan, 1995; Lewis & Shah, 2015), the effects of teachers' use of intentional strategies to promote equitable learning on students' own reports of their experience has not been a central focus of scholarship. We set out to address this gap, because we conjectured that student experience data might serve as a novel and powerful form of formative assessment data for teachers about their own teaching that they could use to inform their instructional decision making and promote more equitable participation in science learning. Our goal and that of the teachers with whom we partnered in this study was to work toward promoting equality of participation across genders and race and ethnicity, and, where possible, to encourage teachers to center the experiences of racially minoritized students and those of girls and gender nonbinary students.

In the approach used in this study, we relied on a novel system for collecting and visualizing student experience data, the Science SEET system (https://www.scienceseet.com/), coupled with a cycle of inquiry we call the Student Experience Improvement Cycle (SEIC), to support teachers in collecting, interpreting, and using student data to improve the quality and equity of students' experience in their classroom. The student experience surveys included in the Science SEET focus on three key aspects of student experience: coherence, relevance, and contribution. The SEIC is a cycle that is based on a Plan-Do-Study-Act cycle (Bryk et al., 2015) that engaged teachers in adapting evidence-based strategies for promoting more equitable small group and whole class knowledge building activities, and then gathering and interpreting student experience data to assess whether the strategy helped improve the overall student experience and differences in experience associated with students' race and gender (see Figure 7).

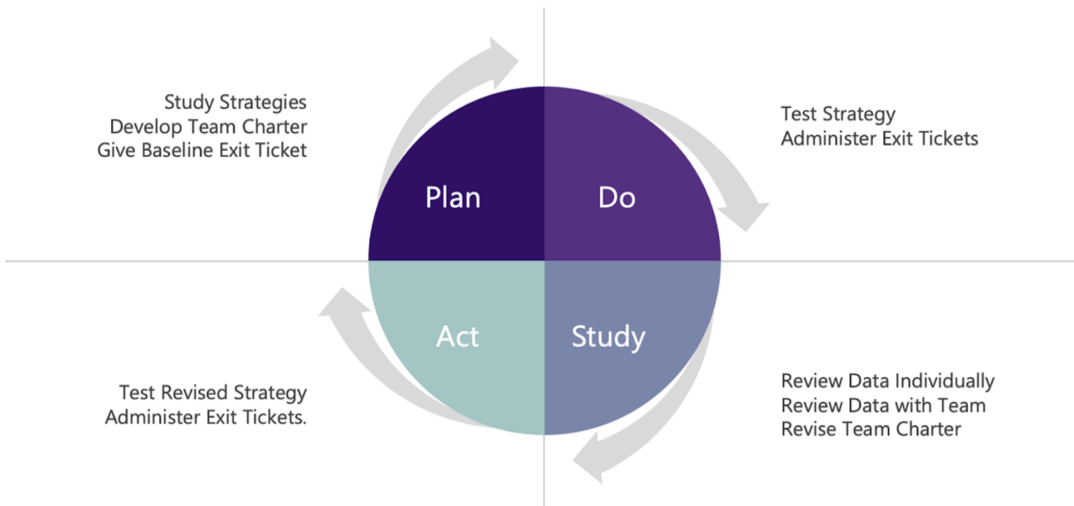


Figure 7: Graphical representation of SEIC

Our approach builds on past research in mathematics that has documented how collaborative communities of teachers can help with improving instruction planning and reflecting on practice together (e.g., Borko et al., 2015). Further, there is evidence that collaboration among mathematics teachers and coaches in inquiry cycles like these involving the use of teacher-collected data can lead to improvement in teaching at scale in mathematics classrooms (Russell et al., 2020). Our work expands this past research by putting students' experiences at the center of improvement and by focusing on science.

Below, we describe the key aspects of experience we aimed to improve for students and the strategies we introduced to teachers to adapt for purposes of improving the equity of student experience.

Strategies Teachers Adapted for Supporting Coherence in Classroom

Reiser and colleagues (2017) argue for the importance of lesson coherence from the students' perspective as necessary to help build a K-12 science classroom environment focused on partnering with students. In their view, "this coherence arises when students see their science work as making progress on questions and problems their classroom community has committed to address, rather than simply following directions from textbooks or teachers" (Reiser et al., 2021, p. 805). In classrooms where students experience coherence, when asked, "Why are you working on this?" students' answers should reflect "what they are figuring out and why it matters to them" (Reiser et al., 2021, p. 807). Such coherence supports students in developing a strong sense of agency over the learning goals in the classroom and has the potential to help sustain interest within classrooms where instruction is organized around investigation of problems and phenomena (Edelson, 2001).

To support students in building knowledge from their questions, several conditions need to be met. First, coherent curriculum materials are needed that support elicitation and use of student questions to drive learning forward. One means to do so that we introduced as a strategy was use of a "Driving Question Board" (Weizman et al., 2008) for capturing students' questions that can organize a sequence of investigations in a unit. A second key resource to support coherence from the student point of view is a classroom routine in which the teacher and students consider what they have figured out and where they need to go next to develop understanding of an anchoring phenomenon or design challenge. This routine, the Navigation Routine (Reiser et al., 2021), supports building connections across lessons and helps students see how the day's lesson fits into the bigger picture of what phenomenon the class is trying to explain or the problem they are trying to solve.

Strategies Teachers Adapted for Improving the Relevance of Activities to Students

According to *A Framework for K-12 Science Education* (National Research Council, 2012), science education should provide students a systemic background to investigate the scientific phenomena related to their "individual and community level priorities" (p. 278). Tan and Calabrese Barton (2012) argue that to support equity in science and mathematics education, there is a need to foster deep connections between the scientific material they are studying and their everyday lives and communities can equip them for participation in public life in a just and democratic fashion. Enhancing the relevance of science can have direct benefits for learning, by increasing motivation and helping students make practical use of science in their everyday lives (Corwin et al., 2015; Penuel, 2016).

There are several evidence-based strategies for enhancing the relevance of science instruction. Some of these involve what Walkington and Bernacki (2014) refer to as group personalization strategies, that is, strategies focused on linking instruction to what a class as a whole might enjoy and find interesting or relate to their values. Strategies introduced to support group personalization introduced include interest surveys to guide the selection of phenomena for instruction (Penuel, Allen, et al., 2022) and engaging students in eliciting and working with phenomena that are related to those selected by the teacher but introduced by students (Reiser et al., 2021). We introduced more individual personalization strategies as well that focused on empowering the use of students' everyday language and heritage languages first in making sense of ideas (Bang et al., 2017; Brown, 2019; González-Howard et al., 2021; Suárez, 2020) . We also supported educators in trying strategies to help students make connections between the day's lesson and their own interests and values (Hulleman & Harackiewicz, 2009).

Strategies Teachers Adapted for Promoting More Equitable Contributions to Knowledge Building

Students' contributions in science classrooms as partners in knowledge building are critical in contemporary science reform (Schwarz et al., 2017), shifting from a model where the teacher is the sole instructional, knowledgeable, and authoring agent. Stroupe (2014) called for students to be involved in science classrooms as epistemic agents, that is, "individuals or groups who take, or are granted, responsibility for shaping the knowledge and practice of a community" (p. 488). A key facet of epistemic agency in contemporary reform is that it is enacted in the context of classroom communities where students are accountable to standards of knowledge development developed by that community (Michaels et al., 2008). Gender and race can play a role in whose contributions are valued in knowledge-building activities, creating patterns of inequity of participation (Carlone, 2004).

To support equitable contribution in knowledge building activities from students' perspective, we introduced several different strategies for teachers to adapt. One was the use of "talk moves," specific utterances that teachers can use to help students share, clarify, or expand their thinking, to help students listen carefully to one another, and helping students think with other students (Michaels & O'Connor, 2015). We also introduced teachers to a strategy to challenge status dynamics by recognizing and affirming competence of learners who are not seen as "smart" by others within small groups and whole-group discussions (Patterson, 2019; Theobald et al., 2017). We also provided strategies for teachers related to how to listen to student ideas: (1) wait time after asking a question for students to be able to think (Rowe, 1974); (2) think-pair-share strategies (Lyman, 1981; Prahl, 2017); and (3) the delaying of evaluation of ideas (Penuel et al., 2017).

Table 6 summarizes key dimensions (constructs) of student experience we analyzed, along with strategies we promoted within the SEIC.

Table 6: Constructs and Strategies for Promoting Equity

Construct	Example Strategies
Coherence Students' understanding of how the day's lesson fits into the broader phenomenon they are studying	Routines for generating questions to guide inquiry Routines for revisiting questions and setting directions
Relevance How relevant students see the day's lesson to themselves and to people in their city or town	Empowering use of everyday language and heritage languages first in making sense of ideas Encouraging students to relate community knowledge and practices to scientific sensemaking Supporting students in using their home languages and life experiences and knowledge to explore the meaning of phenomena
Contribution How students contribute to and benefit from others' contributions to knowledge building in the class.	"Talk Moves" Wait/think time Think-Pair-Share Protocols for hearing from all students (turn-taking strategies) Group students heterogeneously by perceived ability Students co-create and review norms and agreements for group work Amplify the ideas of low-status students for others to work with

Method

This is a design-based research study (Design-Based Research Collective, 2003) in which the professional learning series and research was co-led by the researchers, with assistance from three teacher participants for part of the study. In the study, we set out to address two questions:

1. How does student experience vary by lesson, student, and teacher?
2. Can a professional learning cycle support teachers to use student experience data formatively collaboratively as part of the cycle of inquiry to increase equity with respect to students' experiences of and contributions to knowledge-building activities in the classroom?

Positionalities and Stances of the Leaders and Authors

An important consideration of this line of work is the positionalities of the leaders of the SEIC experience and of the authors. The overall leader of the SEIC was a male-identified, first generation, Pakistani graduate student (first author), who was supported in spring by one white female graduate student, a Latina student research assistant, and the overall PI of the project, a white male (one of the co-equal second authors). In the fall, the project was again led by the same graduate student, but the white graduate student was replaced by a female-identified Vietnamese graduate student. In addition, we were supported directly by three teachers from the spring cohort, all women. One of these teachers was Latina, and the other two identified as white. The other second author is Chinese and the fourth author is a white male. While our group partly mirrored that of the teachers, we also recognize that our own positionalities limit our ability to interpret the experiences of nonbinary and Black students in particular. In addition, we worked directly with teachers, and not with students. For these reasons, we have chosen to limit our interpretation of findings to observable differences in cohorts as to their curriculum and relate those to others' work and to minimize inferences about what may have been going on for students.

The two authors involved in co-leading the SEIC cycles with educators introduced their own stance toward equity to educators as part of each cycle. In the first session, we presented our view that equity demanded centering the experiences of groups of students who had been and continue to be excluded from STEM fields. We asked for agreement among educators to adopt the view that we should look to classroom practices, rather than to deficits in students, to explain patterns in classroom experience data. We provided templates for reflection to encourage this stance as well. And as we reviewed data with educators, we invited them to consider how broader structures in society beyond their own practice might shape and explain patterns observed in their classrooms.

Participants

Participants were recruited in two successive cohorts, with different criteria for inclusion. The first cohort of participants were recruited in early 2021 for a spring professional learning cycle that was intended to focus on remote instruction in science, given the ongoing struggles of teachers in the pandemic to support student participation in science learning. We sought to include 60 teachers in the cohort, and more than 500 applied. We selected participants using a block randomization technique, prioritizing region and race/ethnicity. The second cohort of participants was recruited in early fall 2021 for a fall cycle. This cohort we intended to focus on teachers who had NGSS-aligned curriculum materials, which we thought might better support teachers in making use of student experience data. Specifically, we anticipated recommending strategies be integrated into regularly occurring curricular structures (e.g., DeBarger et al., 2010), and by beginning with knowledge of those structures, we thought we could better support teachers in the improvement process. We sought to include 50 teachers in this cohort, and more than 100 applied. Similar to the first cohort selection, we used block randomization technique but prioritizing teachers' using OpenSciEd and Amplify curricula materials.

As Table 7 below shows, the two cohorts were similar with respect to gender and overall experience teaching. However, the spring cohort was much more racially and ethnically diverse than the fall cohort. We conjecture that the reasons for the difference are due to two factors. First, there were many more applicants than available spots for participants in the spring cohort, and so we were able to ensure greater diversity in the cohort. Second, in purposefully narrowing who could be part of the series in fall to participants using NGSS-aligned materials, we likely limited the diversity of the pool of teachers.

Table 7: Teacher Participants in the Student Experience Improvement Cycle

	Spring 2021	Fall 2021
Gender		
Female	15	16
Male	3	4
Gender nonbinary	0	0
Race/Ethnicity		
white	10	20
Black	1	0
Latinx	2	1
Asian/Asian American	4	1
Native American/Alaskan	3	0

The additional requirements for the fall cohort also helps explain the differences in the types of curriculum materials used (Table 8). The plurality of teachers from the spring cohort used materials they had developed themselves. By contrast, all teachers in the fall used one of two sets of instructional materials: Amplify or OpenSciEd.

Table 8: Curriculum Materials Used by Participants

	Spring 2021	Fall 2021
Amplify	2	6
FOSS	1	0
IQWST	2	0
OpenSciEd	1	14
STEMscopes	1	0
Textbook	6	0
Self-created	5	0

Sources of Data

The source of data for the study was student experience data from “Student Electronic Exit Tickets” or SEETs. Each of the participating teachers collected data about their students' experience using an exit ticket using the Science SEET system.

When registering in the SEET system, students receive an anonymous identifier to use to log in. They are asked to choose from several possible racial and ethnic identities and genders, and they were able to choose more than one identity. For race and ethnicity, we chose to use a slightly modified version of categories for race and ethnicity used by the United States Census. We did so for two reasons. First, we wanted to ensure that groups would not be so small as to be able to describe patterns of their experience (Castillo & Gillborn, 2022). Second, these are categories used in major funding decisions for science education (e.g., NSF) to define groups that are underrepresented in science. For gender, to limit the need to re-code self-identification data, we provided three options: male, female, and gender non-binary. We recognize that our categorization scheme put constraints on ways that students might identify themselves, and that these are not natural categories or "social addresses." We have sought as we describe in analyses below to emphasize variation alongside means for groups, to highlight ways experience differs within groups.

Exit tickets were administered at the conclusion of a lesson and took between three and four minutes to complete. The exit ticket, referred to within the project as the SEET, includes a set of items focused on three dimensions of experience: coherence, relevance, and contribution. The Table 9 below explains each of the dimensions along with the example of items that were used in the SEIC cycles from spring and fall 2021. For these items, students choose either yes, no, or maybe. The SEET system returns visualizations focused on percent yes for teachers to review and that we used in our analysis.

Table 9: SEET Dimensions and Examples of Items

Coherence	Relevance	Contribution
We work together to determine what ideas are most persuasive.	Today's science lesson was personally meaningful.	The teacher encourages us to build on and critique one another's ideas.
The teacher guides us to share our prior experiences or ideas about a phenomenon or topic to inform what we will do next.	I found today's lesson interesting.	Everyone's ideas are heard.
Today we started class by reminding ourselves what we learned in the last class.	If people in my city or town understood the science we learned in today's lesson, they would do something that could help make our city or town a better place.	Did you share any ideas out loud today to the whole class, a small group or a partner?
		If you answered yes to the last question, did any of your ideas influence the class or help others?

The SEET is intended to be used as a "practical measure" of instruction (Yeager et al., 2013). As a practical measure, the purpose is not to develop a reliable scale, but rather a set of items that are good indices of a phenomenon that are practical to administer during a short period of time, over multiple occasions. Validity evidence gathered as part of earlier studies found that (1) the SEET items were readily interpretable by students in ways that corresponded to researchers' intentions; (2) the SEET was easy to administer in a short period of time at the end of class; and (3) SEET responses correlated with performance on a three-dimensional assessment of student learning (Penuel et al., 2018).

Approach to Analysis

We approached the analysis of outcomes using a multi-level modeling framework (Raudenbush & Bryk, 2002). The main outcome of interest was growth with respect to students' experience of coherence, relevance, and contribution. Fitting multilevel models to these data is appropriate, since student observations (completion of a single exit ticket) were nested within students, and students are nested within teachers. For purposes of our analyses, we did not consider differences between classrooms of the same teacher, as most teachers collected data from only one classroom as directed by the researchers.

We also approached the analysis mindful of the ways that in studies such as ours, it is relatively easy to center the perspective of dominant groups. Drawing inspiration from QuantCrit scholars (e.g., Gillborn et al., 2018), we chose to de-center these groups in the choice of reference groups for models. We also note that our analyses were guided by our commitments to promote equity of participation, rather than by a "neutral" stance toward data.

As a first step in analysis, for each outcome construct of interest, we computed an average score based on the responses of each student for occasion q. To model growth over time, we assigned a variable *DAY*, which refers to the number of days since the first data point was collected for a given teacher. For both cohorts, the temporal variable was recentered around the baseline day at which measurements were made to improve interpretation. For example, the baseline day was the first day that data was collected by anyone in the cohort. In the spring, this day was "Day 28" or January 28. In fall, it was "Day 298", or October 25.

To answer the question of whether student experience became more equitable over time, we assigned variables for gender and race/ethnicity at the student level. Categorical variables with more than two responses were dummy coded into multiple binary variables (indicators). For gender, we created three categorical variables from students' self-identification: male (values = 1, 0), female (1, 0), and gender nonbinary (1, 0). For race/ethnicity, we coded students' self-identification into six variables for spring: African American/Black (values = 1, 0), Asian (1, 0), Hispanic/Latinx (1, 0), Native Hawaiian/Pacific Islander (1, 0), Native American/Alaska Native (1, 0), or white (1, 0), including an additional variable for fall: Other (1, 0). Students who selected all race/ethnicity options were considered as missing data, because we interpreted this response as indicative of their resistance to being asked about their race/ethnicity.

A series of hierarchical linear models (HLM) were next fit to the data using HLM 8.2 software (Raudenbush & Bryk, 2021), where *DAY* (level 1) was nested within students (level 2) nested within classrooms (level 3). Specifically, we started with an unconditional means model, which allowed us to quantify variation associated with each level for each of the three outcome variables. We then fit an unconditional growth model to the data by including the temporal predictor, *DAY*, in the models. Next, we report successive conditional growth models for gender and for race/ethnicity, if *DAY* proved to be a significant predictor in the unconditional growth model. We did so first using a reference group (Latinx, female) and then fit separate models for each race/ethnicity and gender. When gender or race/ethnicity predicted growth, we also fit models to the data that included variables for both gender and race/ethnicity. The default full maximum likelihood estimation was applied.

Unconditional Growth Model Specification

The level 1 equation for the unconditional growth models fit to the data was defined as follows:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}*(DAY_{tij}) + e_{tij} \qquad (1)$$

where Y_{tij} denoted the continuous outcome at day t (t=0,1,...,t) for individual i (i=1, 2,...,n) in classroom j (j=0,1,..,j); π_{0ij} was the initial status for student i in classroom j, that is, the expected outcome for student i in classroom j at the baseline timepoint; π_{1ij} was the growth rate for student *i* in classroom *j*; DAY_{tij} was the temporal predictor at time *t* for student *i* in classroom *j*; and e_{tij} was the residual associated with student's score at a specific time point, which was assumed to be normally distributed with a mean of 0 and variance of δ^2.

The unconditional models for the intercept and slope were:

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij} \qquad (2)$$

An important aspect here was that the individually varying parameters that defined the growth trajectory (e.g., the intercept and slope) were assumed to be random.

Conditional Models: Gender

The conditional level 2 equations were formulated by using level 1 intercept and slope as outcomes.

Considering FEMALE as the reference group, the model for estimating effects of gender on growth was defined as:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}*(MALE_{ij}) + \beta_{02j}*(GENDERNB_{ij}) + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}*(MALE_{ij}) + \beta_{12j}*(GENDERNB_{ij}) + r_{1ij} \qquad (3)$$

where the coefficients β_{00j} and β_{10j} represented the mean initial status and rate of change respectively in classroom *j*, controlling for students' gender; r_{0ij} and r_{1ij} were the deviation of each individual's trajectory parameters from their respective means.

Conditional Models: Race/Ethnicity

We selected LATINX group as a reference group which had the most students' identifying themselves in this group other than the white. The level 2 equations were:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}*(WHITE_{ij}) + \beta_{02j}*(BLACK_{ij}) + \beta_{03j}*(ASIAN_{ij}) + \beta_{04j}*(PACIFIC\ ISLANDER_{ij}) + \beta_{05j}*(NATIVE\ AMERICAN_{ij}) + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}*(WHITE_{ij}) + \beta_{12j}*(BLACK_{ij}) + \beta_{13j}*(ASIAN_{ij}) + \beta_{14j}*(PACIFIC\ ISLANDER_{ij}) + \beta_{15j}*(NATIVE\ AMERICAN_{ij}) + r_{1ij} \quad (4)$$

Since the fall cohort has an additional race/ethnicity group, OTHER, comprising approximately 7.5% of the participants, we included it in the fall conditional race/ethnicity model.

Results

The final dataset contained 843 students nested within 45 classrooms for the spring cohort and 1,094 students nested within 48 classrooms for the fall cohort. The detailed descriptive statistics for both cohorts were displayed in Table 10.

Table 10: Descriptive Statistics by Cohort

Variable	Spring 2021	Fall 2021
Outcome Variable	n = 1801 observations	n = 2572 observations
Coherence	M = .77 SD = .26	M = .76 SD = .29
Relevance	M = .50 SD = .32	M = .38 SD = .34
Contribution	M = .58 SD = .28	M = .58 SD = .30
Demographics	n = 843	n = 1094
Gender		
Male	46.25%	45.45%
Female	51.25%	49.22%
Gender nonbinary	2.50%	5.33%
Race/Ethnicity		
white	50.47%	60.15%
Black/African American	22.77%	8.36%
Latinx/Hispanic	11.60%	22.82%
Asian	22.60%	8.32%

Native Hawaiian/Pacific Islander	2.70%	1.09%
American Indian/Alaska Native	5.39%	2.64%
Other	n/a	7.50%

Note. The percent of race is above 100 due to students' selecting more than one race at times.

How Student Experience Varied by Lesson, Student, and Classroom

To address our first research question, we first calculated the variance associated with each level in the unconditional models fit to the data. We then calculated the proportion of variance associated with each level in the model. Results appear below in Table 11.

Table 11: Decomposition of Variance: All Constructs, Spring and Fall Cohort

	Spring Coherence	Spring Relevance	Spring Contribution	Fall Coherence	Fall Relevance	Fall Contribution
Occasion	.0344	.0492	.0458	.0479	.0588	.0520
	(.1856)	(.2218)	(.2140)	(.2189)	(.2425)	(.2281)
Individual	.0311***	.0577***	.0275***	.0253***	.0364***	.0198***
	(.1763)	(.2402)	(.1658)	(.1590)	(.1908)	(.1407)
Classroom	.0055***	.0086**	.0047*	.0064***	.0151***	.0095***
	(.0738)	(.0928)	(.0682)	(.0799)	(.1229)	(.0972)
DAY:Individual	.0000**	.0001 ***	.0000	.0000	.0000	.0000
	(.0033)	(.0069)	(.0021)	(.0019)	(.0020)	(.0014)
DAY:Classroom	.0000	.0000**	.0000***	.0000**	.0000**	.0000**
	(.0018)	(.0035)	(.0037)	(.0016)	(.0016)	(.0019)

Note.***p < .001, **p < .01, *p < .05; inside the parentheses are standard errors.

This analysis showed that more than half of the variance across outcomes was at the lesson level. That is to say, differences in student experience of individual lessons – rather than differences among individuals or across classrooms – accounted for the biggest percentage of variation in classroom experiences. That said, between 29 and 40 percent of the variance in reported experience was at the student level, and between 5 and 10 percent of the variance was due to the teacher. The levels of variation at the student and teacher levels were of particular interest to us, particularly as they relate to students' social identities (gender and race/ethnicity) and teachers' capacity to intervene in ways to improve the quality of student experiences. We therefore continued modeling results through the multi-level framework, turning next to examine growth over time.

We also observed negative correlations between baseline and growth rate across all constructs in the spring at both the student level, ranging from -.496 to -.385, and at the classroom level, ranging from -.662 to -.306. That is to say, those who started at lower levels grew more rapidly over the spring. In this cohort, both individuals and classrooms were becoming more "equitable," in that students who perceived their classroom experience to be of lower quality narrowed the gap with those who perceived their experience to be of higher quality. Further, classrooms became more equal: it mattered less what classroom a student was in for the quality of their experience.

In the fall, the estimated correlations between the baseline and growth rates were negative at the classroom levels but positive at the student levels. Note that the positive correlations were negligible (.035) in coherence, weak in relevance (.226) but strong in contribution (.947). This suggested that students who started higher grew faster over time, while classrooms that started lower may have grown at a faster rate. This pattern shows that for individuals, overall classrooms were becoming somewhat less equitable, following the pattern of a "Matthew Effect" (Merton, 1968) whereby individuals with higher quality experiences at the beginning had even better experiences after their teachers participated in the SEIC. By contrast, classrooms became more equal across the sample: there was less variability across classrooms after participating in the SEIC in students' average quality of experience.

Modeling Change in the Quality of Student Experience: Unconditional Models

To answer our second research question, whether participating in a cycle of inquiry could improve the quality of student experiences, particularly for students historically excluded from science, we modeled growth by examining whether the variable *DAY*—a measure of the day students completed an exit ticket after their teacher began participating in the SEIC—was a significant predictor of student experience. While our research design prevents us from answering whether participating in the SEIC caused a change in students' experience, this longitudinal analysis does permit us to answer the question of whether it is possible to support such a change through the SEIC.

As shown in Table 12, all constructs had positive coefficients in the spring, suggesting increases in students' responses of "yes" over the spring cohort, however, only the coherence construct demonstrated a statistically significant growth of .002742 per day of additional "yes" answers to questions from the exit ticket over time in the unconditional growth model, $t(44)=4.253$, $p<.001$. That is, all students would respond positively to one more coherence question over the course of the year.

For the fall cohort, whereas students' responses increased for relevance and coherence, the negative coefficient for contribution seemed to imply a drop in answering "yes". But among all three constructs, only the relevance score changed significantly over time. Students gained an average of .001099 per day during the study, $t(47)=2.351$, $p<.05$. That is, a third of the students would respond positively to one more relevance question over the course of the year.

Table 12: Unconditional Models for Intercept and Slope from Random effects Models

Fixed	Spring			Fall		
Effects	Coherence	Relevance	Contribution	Coherence	Relevance	Contribution
Intercept	.7155***	.4746***	.5483***	.7482***	.3484***	.5829***
	(.0177)	(.0229)	(.0186)	(.0157)	(.0214)	(.0177)
DAY	.0027***	.0013	.0009	.0004	.0011*	-.0004
	(.0006)	(.0010)	(.0009)	(.0004)	(.0005)	(.0005)

Note. ***p < .001, **p < .01, *p < .05; inside the parentheses are standard errors.

Random	Spring			Fall		
Effects	Coherence	Relevance	Contribution	Coherence	Relevance	Contribution
Intercept	.0311***	.0577***	.0275***	.0253***	.0364***	.0198***

	(.1763)	(.2402)	(.1658)	(.1590)	(.1907)	(.1407)
DAY	.0000**	.0001***	.0000	.0000	.0000	.0000
	(.0033)	(.0069)	(.0021)	(.0019)	(.0020)	(.0014)

Note. ***p < .001, **p < .01, *p < .05; inside the parentheses are standard errors.

To identify potential outliers we first estimated models in which the slopes were fixed, preventing outlying cases from being shrunk to the overall trend. We then plotted the fixed vs random residuals. Except for spring contribution, the extreme cases were spread across multiple classrooms and tended to have fewer data points associated with them. In the case of spring contribution, seven of twelve outlier cases came from two teachers, both of whom were working on strategies to promote greater talk (contribution) in class. But because there was no overall explanation that led us to suspect outlier cases were not legitimate data points, we continued to model slopes randomly as we investigated differential effects by gender and race/ethnicity. The fact that these teachers were focused on the strategy that the outcome that showed growth in fact provides some confirming evidence of our decision to do so.

Modeling Change in the Equity of Student Experience

For the two constructs with significant main effects for growth, we further explored whether there was differential growth for student groups excluded from science. That is, we investigated whether boys and gender nonbinary students grew more in their experience than did girls. We also explored whether students of other races and ethnicities grew more than Latinx students, in terms of the quality of their experience of classrooms. We first modeled growth by gender, before modeling race/ethnicity and then fitting models that included both gender and race/ethnicity.

Modeling Equity in Growth in Coherence for Spring Cohort

Table 13 shows results of our models focused on equity of experience. In general, students were growing positively over time, but there were no differences among genders. With respect to race and ethnicity, baseline experience of our reference group, the Latinx students, was of significantly higher quality as measured by the SEET than that of white (γ_{010} = -.124433, $t(743)$= -2.969, p = .003) and Native Hawaiian/Pacific Islander (γ_{040} = -.161443, $t(743)$= -1.961, p = .050) students. In terms of growth rates, white students grew marginally faster (γ_{110} = .003103, $t(743)$= 1.830, p = .068), while Native Hawaiian/Pacific Islander (γ_{140} = .008945, $t(743)$=2.825, p = .005) and Asian (γ_{130} = .003656, $t(743)$= 2.272, p = .023) students' experience grew in quality at significantly faster rates. American Indian/Alaska Native students grew in the quality of their experience more slowly than Latinx students(γ_{150} = -.003782, $t(743)$= -1.862, p = .063), though this was of marginal significance. We argue that each of these results should be treated with caution, considering the relatively large standard errors and the relatively small number of participants in these racial/ethnic groups.

When we modeled effects of race/ethnicity and gender together, we obtained similar results at similar levels of significance. White and Native Hawaiian/Pacific Islander students began at lower levels, but over time, the quality of their experience improved at a faster rate than did Latinx students. While Asian students grew at a slightly faster rate, American Indian/Alaska Native students grew more slowly. There were, however, no significant gender or combined effects.

Space left blank intentionally for Table 13

Table 13: Modeling Changes to Equity of Experience: Coherence, Spring Cohort

Fixed Effects	Model 1: Unconditional	Model 2: Gender	Model 3: Race/Ethnicity	Model 4: Full Model
Level 1 intercept				
Intercept	.7155*** (.0177)	.7191*** (.0224) Female	.8179*** (.0397) Latinx	.8238*** (.0408) Female Latinx
Male		-.0050 (.0246)		-.0115 (.0252)
Gender non-binary		-.0349 (.0846)		-.0321 (.0802)
white			-.1244** (.0419)	-.1238** (.0413)
Black/African American			-.0587 (.0417)	-.0587 (.0412)
Asian			-.0708 (.0444)	-.0700 (.0440)
Native Hawaiian/Pacific Islander			-.1614 (.0823)†	-.1627* (.0826)
American Indian/Alaska Native			.0572 (.0545)	.0593 (.0535)
Level 1 slope				
Intercept	.0027*** (.0006)	.0028*** (.0008) Female	-.0003 (.0015) Latinx	-.0004 (.0016) Female Latinx
Male		.0001 (.0009)		.0005 (.0010)
Gender non-binary		-.0042 (.0055)		-.0042 (.0052)
white			.0031† (.0017)	.0031† (.0017)
Black/African American			.0012 (.0018)	.0011 (.0018)
Asian			.0037* (.0016)	.0036* (.0016)
Native Hawaiian/Pacific Islander			.0089 ** (.0032)	.0089** (.0031)
American Indian/Alaska Native			-.0038† (.0020)	-.0038* (.0018)

Note. ***p < .001, **p < .01, *p < .05, †p < .10; inside the parentheses are standard errors. Variance of level 1, 2 and 3 available upon request.

Modeling Equity in Growth in Relevance for Fall Cohort

For the fall cohort, because student ratings of relevance increased significantly over the course of the intervention, we analyzed whether growth patterns were related to gender, race/ethnicity, and gender and race/ethnicity together (see Table 14). Notably, ratings began low for this construct. The average growth rate in fall relevance was significantly positive at γ_{100} =.001099, $t(47)=2.351$, $p<.05$. This rate corresponds to students who would say "yes" for one additional relevance item (out of three) over the course of the year. In terms of gender, we found a borderline significant (at alpha = 0.1) negative growth rate for non-binary students (γ_{120} = -.003196, $t(994)$ = -1.708, p = .088), suggesting they showed slower progress than female students over time.

In terms of race/ethnicity, there were no significant differences at baseline, except that white students tended to score slightly lower, γ_{010} = -.056387, $t(986)$ = -1.800, p = .072. Black/African American students' growth in relevance in the fall cohort was significantly greater than Latinx (reference group) students' growth (γ_{120} = .003693, $t(986)$ = 2.671, p = .008). Over the course of a year, Black/African American students would likely increase by one more positive answer than Latinx students would.

In our final model of growth in relevance, which included both gender and race/ethnicity covariates, there were some evidence for the borderline significance in terms of baseline: male students started slightly higher, γ_{010} = .028506, $t(982)$ =

1.684, p = .093, while white students started marginally lower, γ_{030} = -.053147, $t(982)$ = -1.683, p = .093, than the reference group. However, there was no longer significant growth effect of gender. In terms of growth rate, the only statistically significant result was that being Black/African American related positively to individual growth rate, γ_{140} = .003632, $t(982)$ = 2.600, p = .009.

Space left blank intentionally for Table 14

Table 14: Modeling Changes to Equity of Experience: Relevance, Fall Cohort

Fixed Effects	Model 1: Unconditional	Model 2: Gender	Model 3: Race/Ethnicity	Model 4: Full Model
Level 1 intercept				
Intercept	.3484*** (.0214)	.3318*** (.0229) Female	.3869*** (.0334) Latinx	.3689*** (.0335) Female Latinx
Male		.0408† (.0228)		.0385† (.0229)
Gender non-binary		-.0285 (.0481)		-.0323 (.0505)
white			-.0564† (.0313)	-.0531† (.0316)
Black/African American			-.0115 (.0408)	-.0078 (.0412)
Asian			-.0106 (.0447)	-.0089 (.0438)
Native Hawaiian/Pacific Islander			-.0008 (.0776)	-.0085 (.0734)
American Indian/Alaska Native			.0307 (.0788)	.0326 (.0815)
Other			-.0626 (.0408)	-.0605 (.0412)
Level 1 slope				
Intercept	.0011* (.0005)	.0010† (.0005) Female	.0000 (.0010) Latinx	-.0001 (.0010) Female Latinx
Male		.0006 (.0006)		.0006 (.0006)
Gender non-binary		-.0032† (.0019)		-.0030 (.0019)
white			.0010 (.0010)	.0010 (.0010)
Black/African American			.0037*** (.0014)	.0036** (.0014)
Asian			.0010 (.0013)	.0009 (.0014)
Native Hawaiian/Pacific Islander			.0009 (.0021)	.0010 (.0021)
American Indian/Alaska Native			-.0005 (.0032)	-.0004 (.0033)
Other			.0018 (.0018)	.0017 (.0018)

Note. ***p < .001, **p < .01, *p < .05, †p < .10; inside the parentheses are standard errors. Variance of level 1, 2 and 3 available upon request.

We also conducted the unconditional growth model separately for each gender and race/ethnicity group to estimate the fixed and random effects without mixing data across groups. This allows us to focus on the unique experiences of under-represented groups and also to give attention to both average experience and variation within groups. For the fixed effects for spring coherence, by gender, male and female students had a significant and positive growth rate whereas for non-binary students there were too few classrooms that can generate a unique slope for the non-binary students. We also observed little more variance among Non-binary individuals than among other groups, the estimated growth rate is .0001735 and the variance of growth rate was .00020.

By race/ethnicity, we found a significant growth rate for three groups: white, Asian, Native Hawaain/Pacific Islanders in spring coherence. For the fall relevance, we found a significant growth rate for only male students. By race, we found a significant growth for Black or African American students (p=.018), and borderline significant for Native Hawwain and Pacific Islander (p=.096), white students (p=.088).

For the random effects (see Appendix A), in terms of gender, we found that variation in intercept was statistically significant (p<.001) for male and for female students, indicating large variation in their baseline scores. For fall relevance, the variation in individual intercepts was statistically significant for each gender category. Due to limited degree of freedom of nonbinary students in spring, we cannot evaluate the statistical significance of the variation in growth (there were not ample classrooms with enough non-binary students to estimate unique growth terms to evaluate the variance of the growth terms). In terms of race/ethnicity, there was little evidence of variation in slopes or intercepts across groups. The spring coherence scores had significant variance in slope for students belonging to Native Hawaiian and Pacific Islander, but we should interpret this finding with caution due to the limited sample size of this race/ethnicity group. And for fall relevance, the variance in the slope for Latinx or Hispanic (p<.001) and Native Hawaiian and Pacific Islander students was statistically significant (p<.05), indicating wide variability among these groups in how much the quality of their experience changed.

We also conducted separate conditional growth models, by gender and race/ethnicity. For the fixed effects for spring coherence, we did not see any differences in modeling the growth with this approach as compared to conditional growth models in Table 13 other than the Native Hawaiian/Pacific Islanders was more significant (p<.001). Although, there were just 28 students' growth included in models. For fall relevance scores, we did not see any reportable changes in baseline or growth as compared to the conditional growth models in Table 14.

Discussion

In this paper, we investigated how teachers collaboratively used student experience data related to knowledge building activities in science classrooms. During both cohorts, we see encouraging results in improvement in the quality of student experience over time. In the spring cohort, we observed significant and sustained improvement in coherence and for the fall cohort, we saw an increase in students' perceptions of the relevance of lessons after implementing strategies for supporting equitable contribution.

Our results are consistent with past research on cycles of inquiry in disciplinary learning that show that improvements in practice can be achieved (e.g., Russell et al., 2020). What is new here is the focus on the quality of students' experience and on attempting to document how, if at all, inquiry can result in improvements to student experiences in the classroom. Our results provided support for the claim that a Plan-Do-Study-Act cycle can enhance the quality of student experience over the course of a few sessions of professional learning for educators. We view this finding as significant for the field, as the quality of student experience has been shown to be linked to gains in student learning in past large-scale studies (see, e.g., Kane & Staiger, 2012).

Not all aspects of experience improved, however, and only some were statistically significant. In the spring cohort, improvements were primarily to the degree to which students experienced their learning as coherent from their point of view, and in the fall cohort, students' perceived relevance improved. We attribute the difference in findings in part to the instructional materials used in each cohort. In the spring cohort, the fact that many teachers were using their own self-assembled materials meant that coherence was likely to be low, as compared to the phenomenon-based instructional materials that were designed to be coherent from the student perspective (see, e.g., Reiser et al., 2017). The strategies introduced as part of the SEIC helped infuse some elements of coherence designed into other materials, such as using a routine whereby the teacher engaged students in recalling what they did the previous day and helping them to relate the purpose of the day's lesson to a larger arc of learning. In the fall cohort, we were not surprised to see educators focus on relevance, since one of the challenges educators report with phenomenon-based instructional materials is sustaining student interest over multiple weeks of investigating the same phenomenon (Penuel, Reiser, et al., 2022). Further, because the materials were already designed to be coherent and support students' contribution, we did not expect as much potential for growth in those phenomena. At the same time, for both the spring and fall cohorts, we expected more attention—and more growth—on improving students' contributions to knowledge building, but we only found evidence of growth in a few outlier classrooms.

Our findings with respect to whether this intervention has the potential to improve equity of experience were also mixed. On the one hand, we saw Black or African American students' perceptions of relevance increase in the fall cohort. In addition, for the spring cohort, we saw perceptions of coherence among Native Hawaiian/Pacific Islander and Asian students increase. Other groups showed no evidence of a pattern of "reparative" equity. Further, the quality of gender nonbinary students' experience was consistently lower—if not significantly so—across both the spring and fall cohorts. It is possible this particular gap went unnoticed by teachers in their individual data, since teachers are not able to view the experience data disaggregated in the Science SEET system for any group that had less than two individuals in it (Raza et al., 2021). This underscores a theme within the emerging area of collaborative learning analytics (Schneider et al., 2021; Wise et al., 2021) about the challenge of designing tools that meet the needs of its users while also protecting privacy. It may be that such data are valuable at the network or system level, for fostering improvements of the experiences of gender nonbinary students in science classrooms.

This research also contributes to the area of data use in education, particularly with respect to equity. Data use is an extraordinarily common practice in education today, facilitated and fueled by the increasing availability of accountability testing data (Mandinach et al., 2008). But studies of data use in education have pointed to the risk of data use in exacerbating inequities, often by allowing for educators to give voice to and reinforce deficit conceptions of students and their communities (Bertrand & Marsh, 2021; Datnow et al., 2017). In science, studies of data use have shown how practices of looking at accountability test data can undermine a focus on student sensemaking and a focus on improving equity (Braaten et al., 2017). The positive results we observed from our study show the potential for a data use intervention to improve equity. We attribute our results to the combination of a focus on student experiences with a stated up front commitment to improving the quality of experience for those students experiences to whom we owe an education debt (Ladson-Billings, 2006) because of their historical exclusion from science. We also attribute the success of the effort to the close alignment of strategies introduced to teachers to research on learning, which is a key condition for successful use of formative assessment to improve outcomes for students (Penuel & Shepard, 2016).

Limitations of the Study

There were a few limitations of the study. For one, the two design and context of the two cohorts were different, so comparisons across cohorts are not easy to make. In the spring, teachers were largely engaged in remote teaching and used a variety of curriculum materials. In the fall, teachers were teaching in person and used only NGSS-aligned materials. Also, in both SEICs, the constantly changing pandemic conditions proved challenging to many teachers. Despite this fact, we had good attendance and persistence for the work. Second, we cannot draw causal conclusions about the efficacy of the intervention, because we did not randomly assign educators to treatment and comparison groups. However, our longitudinal design did allow for us to document growth and variation in growth to inform future iterations of the SEIC, in preparation for such a study. In addition, understanding the variance components helps us to understand where we might target intervention – and to potentially give a greater role to helping educators refine lessons, since there was so much variance at this level. Third, it is difficult to use our data sources to offer explanations for differences in outcomes by group. This presents a danger when discussing results, in that interpreters can rely on often faulty intuitions and possibly deficit conceptions of groups (or even conceptions of "model minorities") to interpret outcomes. Finally, we do not presume that systemic sexism, transphobia, and systemic racism can be addressed only through cycles of inquiry such as these. The wider society as well as the policies and practices of schools play a much bigger role in reinforcing inequitable patterns of participation—and non-participation—in schools. The SEIC is one way that educators, though, can begin to interrogate their own practice and explore what they can do to address patterns of inequity. In addition, our results suggest the promise of the approach.

Conclusion

Our results highlight the promise of a cycle of inquiry for improving the quality of students' experiences in science classrooms, and also for promoting equitable participation that can actually close gaps in the quality of student experience. At the same time, while we saw improvements to specific dimensions of the quality of student experience, we did not see improvements across all domains, nor do we see a "closing of gaps" among all groups historically excluded from science. More work is needed to develop the SEIC to address the particularities of experiences of nonbinary students, as well as those of racially minoritized students. We also conjecture that more time is needed to see significant gains, and so improvement work in SEIC is unlikely to prove beneficial without deep work of forging long-term partnerships with teachers. It may also benefit from the kinds of one-on-one coaching experiences that have been successful in mathematics (Russell et al., 2017).

We urge other scholars, teachers and stakeholders doing improvement work on student experience to be prepared to facilitate sensemaking in ways that allows for diverse perspectives and work with student experience data in a way that keeps the focus on equity. Improvement is a slow and challenging process and encountering unexpectedness is part of the same process due to histories of systemic inequities of sexism, racism, cultural and linguistic disregards that contribute to inequitable student learning experiences. Partnerships in SEIC surface these tensions and encourage finding evidence-based steps using student experience data and chart new imaginative pathways for creating equitable learning environments for all students.

Chapter V: Discussion and Conclusion

Science students can experience inequities in classrooms based on racial/ethnic and gender identities during knowledge-building activities. In this thesis, I, along with the support of my colleagues and participating teachers, aimed to support equitable students' experiences by race and gender. I led the co-design, development, and study of the use of the *Science SEET system* along with a professional learning series, the *Student Experience Improvement Cycle (SEIC),* to support equity of participation in science classrooms.

The three studies completed in this dissertation emerge from the framework shown in Figure 8 below. We gathered middle school science students' experience of the lesson based on the three experience constructs - coherence, relevance, and contribution - using a closed-ended survey at the end of a lesson. The survey data is gathered using the Science SEET system. After data cleaning and processing, it is presented back to the teachers in the form of different visualizations, a product of co-design efforts, as explained in study 1. After this, teachers engage with these visualization resources to notice and interpret patterns or outliers in the student experience data from their classrooms. In order to plan evidence-based interventions with different instructional strategies to address patterns educators notice in their data, teachers partner within small groups of 3-5 to implement the intervention multiple times in the SEIC cycle for over two months. During the cycle, teachers and researchers set the overarching goal of improving students' experience in the classroom so that we could not predict the quality of a student's experience from knowing their race or gender.

This three-article dissertation is comprised of three completed studies on teachers' use of SEET data:

1. Affordances of co-designing visualizations of student experience data.

2. Teachers' use and their perceived utility of the Science SEET system.
3. Teachers collaborating in the SEIC cycle to foster improvements in student experience data.

Next, I briefly explain each study's main results and conclusions.

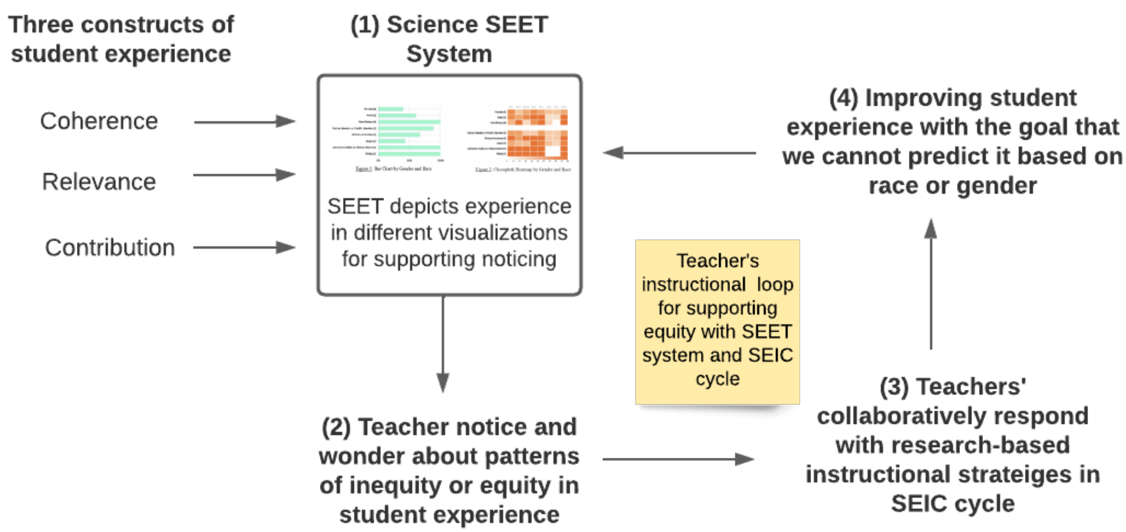


Figure 8: The Framework for this Dissertation

Study 1

We conducted a co-design process to design and select visualizations to prompt teachers noticing related to patterns of equity or inequity in student experience data. We found that different visualizations provide different affordances for reflecting on and supporting equity, and teachers can notice patterns of equity/inequity from the visualizations that emerged as part of the process. We presented teachers' experiences that played a pivotal role in finalizing the visual feedback displays. Further, it inspired innovations in designing different visual feedback displays for the Science SEET system.

The main objective of this study was to provide insights from the design process of the Science SEET system and teachers' thought process. Along with the help of teacher partners in the co-design process, our team designed and developed an exemplar application supporting 'Equity Analytics' (Shah & Reinholtz, 2018). With the wide adoption of tools like ours, teachers can understand inequities of student experiences at scale. Further, our design process showed that we could tailor the system to fit a wide variety of teachers' contexts to facilitate better use and adoption of the tool.

Study 2

The co-design process from study 1 informed the design and development of the Science SEET system, and in this study, we investigated how teachers in a professional development series (the SEIC cycle) used and reflected on the different visualizations of student experience provided in the system. Our study found that visualizations that disaggregated data by race and gender, but not visualizations describing patterns of experience at the classroom level, prompted teachers to notice and reflect on inequities in student participation based on race and gender. However, our study revealed that teachers greatly preferred the 'whole class visualizations' provided through the system to the 'equity visualizations' due to a variety of factors, ranging from their perceived usability of each visualization type, the teachers' own perspectives on equity, and the demographics of their particular classroom.

As such, this study contributes to the field's knowledge and understanding on the utility of learning analytic dashboards for supporting diversity, equity, and inclusion in teaching and learning (Williamson & Kizilcec, 2021; 2022). Our core finding is that the equity visualizations functioned as intended; that is, these three visualizations (Table 2d, 2e, and 2f) prompted teachers to reflect on how student participation patterns varied by race and gender. This finding contrasts with the whole class visualizations, which rarely prompted such equity-oriented reflections.

From a design perspective, however, we need further investigation into new visualization types that embody simplicity and familiarity, while still providing access to finer-grained data on students' participation by race and gender. Equity visualizations necessarily embody more complex data types, suggesting that new visualization literacies may be needed. Future research might focus on how to design learning opportunities for both pre-service and in-service teachers to prepare them to work with visual learning analytics tools. Prior research also suggests that past personal experiences with visualization types can influence the trust people have in them (Peck et al., 2019). Trust is likely an important attribute for visualizations displaying sensitive data by gender and race, raising a question for visualization designers: how does the usability of equity visualizations influence trust in teacher practitioners and other education stakeholders?

Study 3

In this study, we investigated whether the teachers' collaboration in the inquiry cycle (the *SEIC*) can support equity of experience and contributions to knowledge-building activities in the classrooms. We found that the biggest differences in students' experiences were at the lesson level, and not at the individual or classroom level. From the spring teachers' cohort, the coherence construct demonstrated statistically significant growth of "yes" answers to questions from the exit ticket, and for the fall cohort, the relevance score changed significantly over time. We found that some groups' experience improved over time, in ways consistent with our aims for promoting equity of experience, including Native Hawaiian/Pacific Islander (spring coherence), Asian students (spring coherence). For relevance, we found, during the intervention, Black/African American students' growth in relevance in the fall cohort was significantly greater than Latinx (reference group) students' growth. Over the course of a year, Black/African American students would likely give one more positive answer than Latinx students would. There were, however, no significant gender or combined effects in either cohort.

Not all aspects of experience improved, however, and only some observed changes were statistically significant. We attribute the difference in findings from the cohorts in part to the instructional materials used in each cohort. In the spring cohort, the fact that many teachers were using their own self-assembled materials meant that coherence was likely to be low, as compared to the phenomenon-based instructional materials that were designed to be coherent from the student perspective (see, e.g., Reiser et al., 2017). The strategies introduced as part of the SEIC helped infuse some elements of coherence designed into other materials. In the fall cohort, we were not surprised to see educators focus on relevance, since one of the challenges educators report with phenomenon-based instructional materials is sustaining student interest over multiple weeks of investigating the same phenomenon (Penuel, Reiser, et al., 2022).

Our findings with respect to whether this intervention has the potential to improve equity of experience were also mixed. On the one hand, we saw Black or African American students' perceptions of relevance increase in the fall cohort. But the quality of gender nonbinary students' experience was consistently lower—if not significantly so—across both the spring and fall cohorts. It is possible this particular gap went unnoticed by teachers in their individual data, since teachers are not able to view the experience data disaggregated in the Science SEET system for any group that had less than two individuals in it (Raza et al., 2021). This underscores a theme within the emerging area of collaborative learning analytics (Schneider et al., 2021; Wise et al., 2021) about the challenge of designing tools that meet the needs of its users while also protecting privacy. Our results highlight the promise of a cycle of inquiry for improving the quality of students' experiences in science classrooms, and also for promoting equitable participation that can actually close gaps in the quality of student experience.

Limitations

There are some limitations across the studies that are of note. First, SEIC teachers did not share a common district context. This made it difficult in some cases for teachers to learn effectively from one another. Second, we cannot know for sure whether the improvements we saw to the quality of experience were due to the SEIC, because we did not have a comparison group design. The cohorts themselves were quite different as well: teachers used different curriculum, and in the spring cohort, teachers were largely teaching remotely. Third, the Science SEET system does not visualize any student group data by gender or race, if there are fewer than two students, to protect the anonymity of individual students. During the SEIC, some teachers expressed interest in seeing all students' data to better support the equity goals. Without a policy and protocol, we fear using individual student experience data might do more harm than good in helping support equity. However, protecting student privacy might have limited conversations on supporting students whose numbers were small in the classrooms. These include gender non-binary students, whose experience was visible to us as researchers, but rarely to individual teachers. Fourth, the second and third studies were conducted in the COVID-19 pandemic, which created significant challenges for educators and students in the schools due to uncertainty over conditions and unwelcome school closures. That impacted teachers' implementation of the instructional strategies and their collaboration in the SEIC cycle. Fifth, teachers rarely used the system after the conclusion of the professional learning cycle, limiting our opportunities to study its use. Finally, the cycles of the SEIC are immensely time consuming as well. We cannot provide the system to the user and expect to support equitable instruction rather it is supposed to be embedded in the teacher's context. To support adoption of the SEET system and foster collaboration in SEIC, we have to provide deeper facilitation and support to the teachers. One example relates to communicating with the teachers: over the course of the study, I responded to over fifteen hundred emails to teachers in both spring and fall.

Future Work

There are three main future directions that emerge from this dissertation for organizing research for social justice in K12 settings.

In study 3 and retrospective observations in the SEIC, we found that teachers had a difficult time interpreting across different teaching contexts during their collaboration. It was complex to see how the strategies tied up without having a common context related to school or district. We imagine a possible future to run this Science SEET system and SEIC where teachers have a common educational context to allow them greater opportunities for coherent implementation and work together toward improvement toward common system level goals.

In study 2, we found that teachers preferred visualizations of "whole classroom" student experience data that gave them no meaningful data on equity due to their classroom context or ease-of-use in visualization type. In the future, we foresee either limiting the study to classrooms with lots of diversity, or only providing visualizations that disaggregate data by race and gender to support more reflection on equity of participation among teachers.

Our intervention with only one iteration of SEIC with the same teachers provided evidence that improvement in quality of student experience is possible. But we fell short of our equity aim to support the experience of all students in the classrooms by race and gender. We heard from teachers part of the SEIC and our investigation also suggests that multiple cycles to SEIC are needed to yield improvement in equitable growth in experience across all races and genders. Further, much-concentrated work within a school district or research-practice partnership can support more equitable improvement in the student experience. More cycles would have the benefit of getting educators more used to the system itself.

REFERENCES

Achieve. (2016). EQuIP rubric for lessons and units: Science, Version 3.0. Author.

Achieve. (2017). Using phenomena in NGSS-designed lessons and units.

Achieve. (2018). Science task screener. Author.

Albers, D., Correll, M., Gleicher, M., & Franconeri, S. (2014). Ensemble processing of color and shape: Beyond mean judgments. *Journal of Vision, 14*(10), 1056-1056. https://doi.org/10.1167/14.10.1056

Ali, Liaqat, Mohsen Asadi, Dragan Gašević, Jelena Jovanović, and Marek Hatala. "Factors influencing beliefs for adoption of a learning analytics tool: An empirical study." *Computers & Education 62* (2013): 130-148. https://doi.org/10.1016/j.compedu.2012.10.023

Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in context: reaching beyond usability in learning Analytics Dashboard Design. *Journal of Learning Analytics, 6(*2), 70-85. https://doi.org/10.18608/jla.2019.62.5

Alzen, J., Edwards, K., Penuel, W. R., Reiser, B. J., Passmore, C., Griesemer, C., Zivic, A., & Murzynski, C. (2020). Exploring the connections between epistemic agency and a commitment to the collective enterprise of sensemaking in the science classroom. In M. Gresalfi, I. S. Horn, N. Enyedy, H. J. So, V. Hand, K. Jackson, S. E. McKenney, A. Lefstein, & T. M. Philip (Eds.), *Proceedings of the International Conference of the Learning Sciences*. International Society of the Learning Sciences.

Arthars, N., & Liu, D. Y. T. (2020). How and why faculty adopt learning analytics. *In Adoption of Data Analytics in Higher Education Learning and Teaching* (pp. 201-220). Springer, Cham. https://doi.org/10.1007/978-3-030-47392-1_11

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Brown, B. A. (2019). *Science in the city: Culturally relevant STEM education*. Harvard Education Press.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching, 53*(7), 1082-1112. https://doi.org/10.1002/tea.21257

Borko, H., Jacobs, J., Koellner, K., & Swackhamer, L. E. (2015). *Mathematics professional development: Improving teaching using the problem-solving cycle and leadership preparation models*. Teachers College Press.

Bang, M., Brown, B. A., Calabrese Barton, A., Rosebery, A., & Warren, B. (2017). Toward more equitable learning in science: Expanding relationships among students, teachers, and science practices. In C. Schwarz, C. Passmore, & B. J. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 33-58). NSTA.

Bertrand, M., & Marsh, J. A. (2021). How data-driven reform can drive deficit thinking. *Phi Delta Kappan*, 102(8), 35-39. https://doi.org/10.1177/00317217211013936
Braaten, M., Bradford, C., Kirchgasler, K. L., & Barocas, S. F. (2017). How data use for accountability undermines equitable science education. *Journal of Educational Administration, 55*(4), 427-446. https://doi.org/10.1108/JEA-09-2016-0099
Carlone, H. B. (2004). The cultural production of science in reform-based physics: Girls' access, participation, and resistance. *Journal of Research in Science Teaching, 41*(4), 392-414. https://doi.org/10.1002/tea.20006
Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.
Carlone, H. B., Haun-Frank, J., & Webb, A. (2011). Assessing equity beyond knowledge- and skills-based outcomes: A comparative ethnography of two fourth-grade reform-based science classrooms. *Journal of Research in Science Teaching, 48*(5), 459-485. https://doi.org/10.1002/tea.20413
Carlone, H., Scott, C. M., & Lowder, C. (2014). Becoming (less) scientific: A longitudinal study of students' identity work from elementary to middle school science. *Journal of Research in Science Teaching, 51*(7), 836-869. https://doi.org/10.1002/tea.21150
Calabrese Barton, A., & Osborne, M. D. (Eds.). (2001). *Teaching science in diverse settings: Marginalized discourses and classroom practices*. Peter Lang.
Cohen, E. G., & Lotan, R. A. (1995). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal, 32*(1), 99-120. https://doi.org/10.2307/1163215
Corwin, L. A., Runyon, A., Robinson, A., & Dolan, E. L. (2015). The laboratory course assessment survey: a tool to measure three dimensions of research-course design. *CBE—Life Sciences Education, 14*, AR37. https://doi.org/10.1187/cbe.15-03-0073
Castillo, W., & Gillborn, D. (2022). *How to "QuantCrit:" Practices and questions for education data researchers and users*. (EdWorkingPaper No. 22-546). Annenberg Institute at Brown University. https://doi.org/EdWorkingPaper No. 22-546
Digital Promise. 2011. Diversity statement. Retrieved September 23rd, 2022 from https://digitalpromise.org/wp-content/uploads/2016/09/lpsgrowing_diversity_FINAL-1.pdf
Dourado, R. A., Rodrigues, R. L., Ferreira, N., Mello, R. F., Gomes, A. S., & Verbert, K. (2021, April). A teacher-facing learning analytics dashboard for process-oriented feedback in online learning. *In LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 482-489). https://doi.org/10.1145/3448139.3448187
Dazo, S. L., Stepanek, N. R., Chauhan, A., & Dorn, B. (2017, May). Examining instructor use of learning analytics. *In Proceedings of the 2017 CHI Conference*

Extended Abstracts on Human Factors in Computing Systems (pp. 2504-2510). https://doi.org/10.1145/3027063.3053256
Diary Studies: Understanding Long-Term User Behavior and Experiences. Retrieved August 23, 2022 from https://www.nngroup.com/articles/diarystudies/
DeBarger, A. H., Penuel, W. R., Harris, C. J., & Schank, P. (2010). Teaching routines to enhance collaboration using classroom network technology. In F. Pozzi & D. Persico (Eds.), *Techniques for fostering collaboration in online learning communities: Theoretical and practical perspectives* (pp. 222-244). IGI Global.
Datnow, A., Greene, J. C., & Gannon-Slater, N. (2017). Data use for equity: implications for teaching, leadership, and policy: Introduction to data use and equity. *Journal of Educational Administration, 55*(4), 354-360. https://doi.org/10.1108/JEA-04-2017-0040
Eccles, J. S. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75-146). W. H. Freeman.
Engle, R. A. (2012). The productive disciplinary engagement framework: Origins, key concepts, and developments. In Y. Dai (Ed.), *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning* (pp. 161-200). Routledge.
Ericcson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). MIT Press.
Edelson, D. C. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching, 38*(3), 355-385. https://doi.org/10.1002/1098-2736(200103)38:3<355::AID-TEA1010>3.0.CO;2-M
Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de Leon, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: findings from a four-year intervention study. *Instructional Science, 44*(3), 267-291. https://doi.org/doi.org/10.1007/s11251-016-9371-3
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
Greider, C. W., Sheltzer, J. M., Cantalupo, N. C., Copeland, W. B., Dasgupta, N., Hopkins, N., ... & Wong, J. Y. (2019). Increasing gender diversity in the STEM research workforce. *Science, 366*(6466), 692-695. DOI: 10.1126/science.aaz0649
Gruber, J., Mendle, J., Lindquist, K. A., Schmader, T., Clark, L. A., Bliss-Moreau, E., ... & Williams, L. A. (2021). The future of women in psychological science. *Perspectives on Psychological Science, 16*(3), 483-516. DOI: 10.1177/1745691620952789
Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of knowing: Individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19-25. https://doi.org/10.3102/0013189X032005019

Gunckel, K. L., & Tolbert, S. (2018). The imperative to move toward a dimension of care in engineering education. *Journal of Research in Science Teaching, 55*(7), 938-961. https://doi.org/10.1002/tea.21458

González-Howard, M., Andersen, S., & Pérez, K. M. (2021). Enhancing science lessons to support multilingual students' engagement in science and engineering practices. *Science Scope, 44*(3), 24-31.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM, 53*(6), 59-67.

Hulleman, C., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*(1410-1412). https://doi.org/10.1126/science.1177067

Jaber, L. Z., & Hammer, D. (2016). Engaging in science: A feeling for the discipline. *Journal of the Learning Sciences, 25*(2), 156-202. https://doi.org/10.1080/10508406.2015.1088441

Krumm, A. E., Penuel, W. R., Pazera, C., & Landel, C. (2020). Measuring equitable science instruction at scale. In M. Gresalfi, I. S. Horn, N. Enyedy, H.-J. So, V. Hand, K. Jackson, S. E. McKenney, A. Leftstein, & T. M. Philip (Eds.), *Proceedings of the International Conference of the Learning Sciences* (Vol. 4, pp. 2461-2468). International Society of the Learning Sciences.

Krist, C., & Suárez, E. (2018). Doing science with fidelity to persons: Instantiations of caring participation in science practices. In J. Kay & R. Luckin (Eds.), *Proceedings of the 13th International Conference of the Learning Sciences* (Vol. 1, pp. 424-431). International Society of the Learning Sciences.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper: MET Study*. Bill & Melinda Gates Foundation.

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology* (pp. 21-34). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9088-8_2

Li, Q., Jung, Y., & Friend Wise, A. (2021, April). Beyond first encounters with analytics: Questions, techniques and challenges in instructors' sensemaking. *In LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 344-353). https://doi.org/10.1145/3448139.3448172

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459. https://doi.org/10.1177/0002764213479367

Langer-Osuna, J. M. (2015). From getting "fired" to becoming a collaborator: A case of the coconstruction of identity and engagement in a project-based mathematics classroom. *Journal of the Learning Sciences, 24*(1), 53-92. https://doi.org/10.1080/10508406.2014.944643

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press.

Lewis, C. M., & Shah, N. (2015). How equity and inequity can emerge in pair programming. In B. Dorn, J. Sherard, & Q. Cutts (Eds.), *ICER '15: Proceedings* *of the eleventh annual International Conference on International Computing Education Research* (pp. 41-50). Association for Computing Machinery. https://doi.org/10.1145/2787622.2787716

Lyman, F. T. (1981). The responsive classroom discussion: The inclusion of all students. In A. S. Anderson (Ed.), *Mainstreaming Digest* (pp. 109-113). University of Maryland Press.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher, 35*(7), 3-12. https://doi.org/10.3102/0013189X035007003

Master, A., & Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology, 12*(1), 152-198. https://genderandset.open.ac.uk/index.php/genderandset/article/view/674.

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching, 55*(7), 1053-1075. https://doi.org/10.1002/tea.21459

Molenaar, I., & Knoop-van Campen, C. A. (2018). How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies, 12*(3), 347-355. https://doi.org/10.1109/TLT.2018.2851585

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. sage.

Munter, C., & Haines, C. (2019). Mathematics teachers' enactment of cognitively demanding tasks and students' perception of racial differences in opportunity. *Mathematical Thinking and Learning, 21*(3), 155-177. https://doi.org/10986065.2019.1576002

Manz, E. (2015). Representing argumentation as functionally emergent from scientific activity. *Review of Educational Research, 85*(4), 553-590. https://doi.org/10.3102/0034654314558490

Manz, E. (2019). Getting a grip: A framework for designing and adapting elementary school science investigations. *Science and Children, 56*(8), 80-87.

Morozov, A., Herrenkohl, L., Shutt, K., Thummaphan, P., Vye, N., Abbott, R. D., & Scalone, G. (2014). Emotional engagement in agentive science environments. In J. L. Polman, E. Kyza, K. O'Neill, & I. Tabak (Eds.), *Proceedings of the 11th International Conference of the Learning Sciences* (pp. 1152-1156). International Society of the Learning Sciences.

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in the Philosophy of Education, 27*, 283-297. https://doi.org/10.1007/s11217-007-9071-1

Merton, R. K. (1968). The Matthew Effect in science: The reward and communication systems of science are considered. *Science, 159* (3810), 56-63. https://doi.org/10.1126/science.159.3810.56

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussions. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), Socializing intelligence through talk and dialogue (pp. 333-347). *American Educational Research Association.*

Mandinach, E. B., Honey, M., Light, D., & Brunner, C. (2008). A conceptual framework for data-driven decision making. In E. B. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 13-31). Teachers College Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. National Academies Press. https://doi.org/10.17226/18290

Ochoa, X., Knight, S., & Wise, A. F. (2020). *Learning analytics impact: Critical conversations on relevance and social responsibility*. Journal of Learning Analytics, 7(3), 1-5. https://doi.org/10.18608/jla.2020.73.1

O'Neill, T. B. (2010). *Fostering spaces of student ownership in middle school science*. Equity & Excellence in Education, 41(1), 6-20. https://doi.org/10.1080/10665680903484909

OpenSciEd. (2019). OpenSciEd teacher handbook. Retrieved September 25, 2022. https://www.openscied.org/teacher-handbook/

Penuel, W. R., & Watkins, D. A. (2019). Assessment to promote equity and epistemic justice: A use-case of a research-practice partnership in science education. *The ANNALS of the American Academy of Political and Social Science, 683*(1), 201-216.

Penuel, W. R., Van Horne, K., Jacobs, J. J., & Turner, M. (2018). Developing a validity argument for practical measures of student experience in project-based science classrooms. *In Annual Meeting of the American Educational Research Association*, New York, NY.

Penuel, W., Van Horne, K., Severance, S., Quigley, D., & Sumner, T. (2016). Students' responses to curricular activities as indicator of coherence in project-based science. Singapore: *International Society of the Learning Sciences.*

Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of Research on Teaching* (pp. 787-851). AERA.

Penuel, W. R., DeBarger, A. H., Boscardin, C. K., Moorthy, S., Beauvineau, Y., Kennedy, C., & Allison, K. (2017). Investigating science curriculum adaptation as a strategy to improve teaching and learning. *Science Education, 101*(1), 66-98. https://doi.org/10.1002/sce.21249

Penuel, W. R. (2016). Studying science and engineering learning in practice. *Cultural Studies of Science Education, 11*(1), 89-104.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. University of Chicago Press.

Paris, D., & Alim, H. S. (Eds.). (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press.

Project for Education Research That Scales (PERTS). 2010. Retrieved September 15th, 2022 from https://www.perts.net/

Peck, E. M., Ayuso, S. E., & El-Etr, O. (2019, May). Data is personal: Attitudes and perceptions of data visualization in rural Pennsylvania. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12). https://doi.org/10.1145/3290605.3300474

Penuel, W. R., Allen, A.-R., Henson, K., Campanella, M., Patton, R., Rademaker, K., Reed, W., Watkins, D. A., Wingert, K., Reiser, B. J., & Zivic, A. (2022). Learning practical design knowledge through co-designing storyline science curriculum units. *Cognition and Instruction, 40*(1), 148-170. https://doi.org/10.1080/07370008.2021.2010207

Patterson, A. (2019). Equity in groupwork: the social process of creating justice in a science classroom. *Cultural Studies of Science Education, 14*, 361-381. https://doi.org/10.1007/s11422-019-09918-x

Prahl, K. (2017). Best practices for the think-pair-share active-learning technique. *The American Biology Teacher, 79*(1), 3-8. https://doi.org/10.1525/abt.2017.79.1.3

Penuel, W. R., Reiser, B. J., McGill, T. A. W., Novak, M., Van Horne, K., & Orwig, A. (2022). Connecting student interests and questions with science learning goals through project-based storylines. *Disciplinary and Interdisciplinary Science Education Research, 4*(1), 1-27. https://doi.org/10.1186/s43031-021-00040-z

Reinholz, D. L., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education, 49*(2), 140-177.

Reinholz, D., & Shah, N. (2021). Equity and Equality: How Data Visualizations Mediate Teacher Sensemaking About Racial and Gender Inequity. *Contemporary Issues in Technology and Teacher Education, 21*(3), 646-669. https://www.learntechlib.org/primary/p/218443/

Reiser, B. J., Novak, M., & McGill, T. A. (2017). *Coherence from the students' perspective: Why the vision of the framework for K-12 science requires more than simply "combining" three dimensions of science learning*. In Board on Science Education Workshop "Instructional Materials for the Next Generation Science Standards.

Reiser, B. J., Novak, M., McGill, T. A., & Penuel, W. R. (2021). Storyline units: An instructional model to support coherence from the students' perspective. *Journal of Science Teacher Education, 32*(7), 805-829. https://doi.org/10.1080/1046560X.2021.1884784

Roth, W. M., & Jornet, A. (2014). Toward a theory of experience. *Science Education, 98* (1), 106-126. https://doi.org/10.1002/sce.21085

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993, May). The cost structure of sensemaking. *In Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 269-276).

Raza, A., Penuel, W. R., Allen, A.-R., Sumner, T., & Jacobs, J. K. (2021). "Making it culturally relevant": A visual learning analytics system supporting teachers to reflect on classroom equity. In E. de Vries, Y. Hod, & J. Ahn (Eds.), *Proceedings of the 15th International Conference of the Learning Sciences* (pp. 442-449). International Society of the Learning Sciences.

Raza, A., Penuel, W. R., Jacobs, J., & Sumner, T. (2020). Supporting equity in schools: Using visual learning analytics to understand students' classroom experiences. In M. Schmidt, A. Tawfik, Y. Earnshaw, & I. Jahnke (Eds.), *Learner and user experience research: An introduction for the field of learning design & technology* (pp. 1-20). Edtechbooks. https://edtechbooks.org/ux/supporting_school_equity

Russell, J. L., Correnti, R. J., Stein, M. K., Bill, V., Hannan, M., Schwartz, N. L., Booker, L. N., & Pratt, N. R. (2020). Learning from adaptation to support instructional improvement at scale: Understanding coach adaptation in the TN Mathematics Coaching Project. *American Educational Research Journal, 57*(1), 148-187. https://doi.org/10.3102/0002831219854050

Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence on language, logic, and fate control: Part one - wait time. *Journal of Research in Science Teaching, 11*, 81-94.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Raudenbush, S. W., & Bryk, A. S. (2021). *HLM 8: Hierarchical linear and nonlinear modeling*. SSI Scientific Software.

Shah, N., Christensen, J. A., Ortiz, N. A., Nguyen, A. K., Byun, S., Stroupe, D., & Reinholz, D. L. (2020). Racial hierarchy and masculine space: Participatory in/equity in computational physics classrooms. *Computer Science Education, 30*(3), 254-278. https://doi.org/10.1080/08993408.2020.1805285

Schwarz, C. V., Passmore, C. M., & Reiser, B. J. (2017). Moving beyond "knowing" science to making sense of the world. In C. V. Schwarz, C. M. Passmore, & B. J. Reiser (Eds.), *Helping students make sense of the world through next generation science and engineering practices* (pp. 3-21). Arlington, VA: NSTA Press.

Stroupe, D. (2014). Examining classroom science practice communities: How teachers and students negotiate epistemic agency and learn scienceas-. practice. *Science Education, 98*(3), 487-516. https://doi.org/10.1002/sce.21112

Shah, N., & Coles, J. A. (2020). Preparing teachers to notice race in classrooms: Contextualizing the competencies of preservice teachers with antiracist inclinations. *Journal of Teacher Education, 71*(5), 584-599. https://doi.org/10.1177/0022487119900204

Stroupe, D. (2021). Naming and disrupting epistemic injustice across curated sites of learning. *Journal of the Learning Sciences, 31*(2), 317-334. https://doi.org/10.1080/10508406.2021.1977647

Suárez, E. (2020). “Estoy Explorando Science”: Emergent bilingual students problematizing electrical phenomena through translanguaging. *Science Education*. https://doi.org/10.1002/sce.21588
Schneider, B., Dowell, N., & Thompson, K. (2021). Collaboration analytics — Current state and potential futures. *Journal of Learning Analytics, 8*(1), 1-12. https://doi.org/10.18608/jla.2021.7447
Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, A. J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PloS one, 12*(7), e0181336. https://doi.org/10.1371/journal.pone.0181336
Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. *Cognition and Instruction, 22*(4), 393-429. https://doi.org/10.1207/s1532690Xci2204_2
Tan, E., & Calabrese Barton, A. (2012). *Empowering science and mathematics education in urban schools.* University of Chicago Press.
Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education, 122*, 119-135. https://doi.org/10.1016/j.compedu.2018.03.018
Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020, March). Learning analytics dashboards: The past, the present and the future. *In Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 35-40). https://doi.org/10.1145/3375462.3375504
Vygotsky, L. S. (1994). The problem of the environment. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp. 338-354). Blackwell.
Wright, C., Wendell, K. B., & Paugh, P. B. (2018). “Just put it together to make no commotion:” re-imagining urban elementary students’ participation in engineering design practices. *International Journal of Education in Mathematics, Science, and Technology, 6*(3), 285-301. https://doi.org/10.18404/ijemst.428192
Warren, B., & Rosebery, A. S. (2011). Navigating interculturality: African American male students and the science classroom. *Journal of African American Males in Education* (JAAME), 2(1), 98-115.
Williamson, K., & Kizilcec, R. (2022, March). A review of learning analytics dashboard research in higher education: Implications for justice, equity, diversity, and inclusion. *In LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 260-270). https://doi.org/10.1145/3506860.3506900
Williamson, K., & Kizilcec, R. F. (2021, June). Learning Analytics Dashboard Research Has Neglected Diversity, Equity and Inclusion. *In Proceedings of the Eighth ACM Conference on Learning@ Scale* (pp. 287-290). https://doi.org/10.1145/3430895.3460160
Wise, A. F., & Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making. *Journal of Learning Analytics, 6*(2), 53-69. https://doi.org/10.18608/jla.2019.62.4

Wise, A. F., Sarmiento, J. P., & Boothe Jr, M. (2021, April). Subversive learning analytics. *In LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 639-645). https://doi.org/10.1145/3448139.3448210

Windschitl, M., & Calabrese Barton, A. (2016). Rigor and equity by design: Seeking a core of practices for the science education community. *In D. H. Gitomer & C. A. Bell (Eds.), Handbook of Research on Teaching* (pp. 787-851). AERA.

Weizman, A., Shwartz, Y., & Fortus, D. (2008). The driving question board: a visual organizer for project-based science. *The Science Teacher, 75*(8), 33-37.

Walkington, C., & Bernacki, M. L. (2014). Motivational interventions as motivating students by "personalizing" learning and individual interests: A consideration of theory, design, and implementation issues. *In Advances in motivation and achievement* (pp. 139-176). Emerald Group Publishing Limited.

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In U. Cress, C. P. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 425-443). Springer. https://doi.org/10.1007/978-3-030-65291-3_23

Zivic, A., Smith, J. F., Reiser, B. J., Edwards, K. D., Novak, M., & McGill, T. A. W. (2018). Negotiating epistemic agency and target learning goals: Supporting coherence from the students' perspective. In J. Kay & R. Luckin (Eds.), *Proceedings of the 13th International Conference of the Learning Sciences* (Vol. 1, pp. 25-32). International Society of the Learning Sciences.

Appendix A

Table A1: Unconditional Models (Random Effects): Spring Coherence and Fall Relevance, Each Gender

Random Effects

	Spring Coherence		Fall Relevance	
	Variance Component	SD	Variance Component	SD
Female				
Occasion	.0340	.1844	.0572	.2392
Individual	.0286***	.1692	.0357***	.1890
Slope	.0000	.0015	.0000†	.0032
Male				
Occasion	.0336	.1834	.0606	.2462
Individual	.0358***	.1892	.0377***	.1943
Slope	.0000**	.0038	.0000	.0004
Gender non-Binary				
Occasion	.0381	.1952	.0443	.2104
Individual	.0624	.2498	.0477**	.2185
Slope	.0002	.0142	.0000	.0010

***p < .001, **p < .01, *p < .05, †p < .10. Statistical significance could not be evaluated because there were too few classrooms that contained a sufficient number of spring non-binary students to estimate unique slopes.

Table A2: Unconditional Models (Random Effects): Spring Coherence and Fall Relevance, Each Race/Ethnicity

Random Effects	Spring Coherence		Fall Relevance	
	Variance Component	SD	Variance Component	SD
Latinx/Hispanic				
Occasion	.0361	.1899	.0480	.2190
Individual	.0237	.1540	.0556***	.2357
Slope	.0000	.0021	.0001***	.0069
white				
Occasion	.0391	.1978	.0578	.2403
Individual	.0321***	.1793	.0355***	.1885
Slope	.0000†	.0024	.0000	.0010
Black/African American				
Occasion	.0388	.1971	.0612	.2473
Individual	.0171**	.1308	.0252†	.1588
Slope	.0000†	.0033	.0000	.0004
Asian				
Occasion	.0264	.1625	.0536	.2315
Individual	.0251***	.1584	.0325**	.1804
Slope	.0000	.0037	.0000	.0014
Native Hawaiian/Pacific Islander				
Occasion	.0565	.2377	.0353	.1880
Individual	.0011*	.0324	.0404**	.2009
Slope	.0000*	.0006	.0000*	.0040
American Indian/Alaska Native				
Occasion	.0530	.2302	.0338	.1838
Individual	.0119	.1089	.0440	.2099
Slope	.0000	.0012	.0000	.0022
Other				
Occasion	n/a	n/a	.0478	.2185
Individual	n/a	n/a	.0394**	.1986
Slope	n/a	n/a	.0000	.0011

***p < .001, **p < .01, *p < .05, †p < .10.

Appendix B

Teacher Data Notebook: Looking at My SEET Data

Data Note: 1
Date:
Data I am looking at:

What feelings are present, as you begin to look at the data? (You can circle words or use any colors to highlight relevant emotion words provided here if you like)

Curious	Mistrustful
Engrossed	Displeased
Stimulated	Angry
Encouraged	Horrified
Proud	Puzzled
Surprised	Numb
Invigorated	Troubled
Thankful	Ashamed
Delighted	Exhausted
Pleased	Hurt
Thrilled	Unhappy
Fulfilled	Anxious
Relieved	Sensitive

What patterns do you notice in the data, and what do you wonder about?

Things I notice:	Things I am wondering about:	Which Visual representation(s) you are looking at for that pattern.